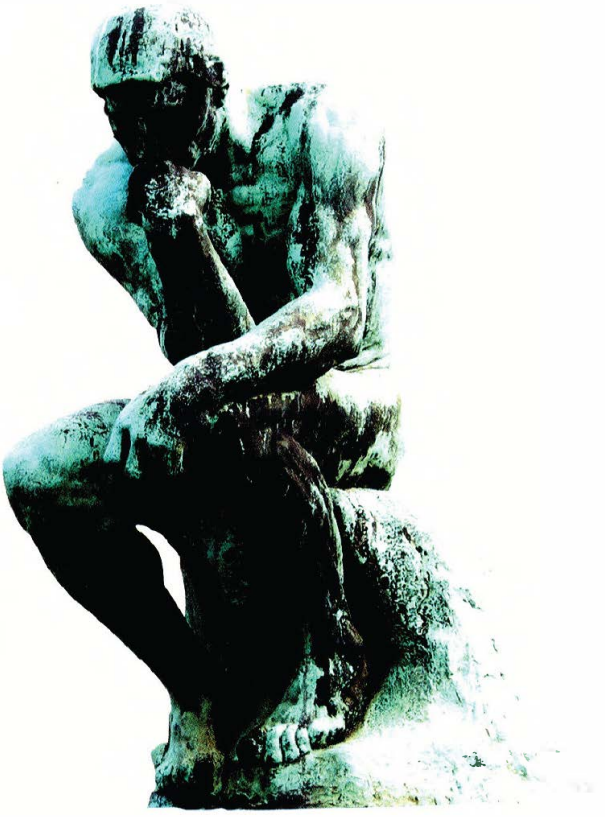
LA ÉTICA Y LOS LÍMITES DE LA FILOSOFÍA

Bernard Williams



8

CÁTEDRA

La ética y los límites de la filosofía

Colección Teorema
Serie mayor

Bernard Williams

La ética y los límites de la filosofía

Prólogo de Jonathan Lear
Comentario de A. W. Moore

Traducción de Sergi Rosell

CÁTEDRA
TEOREMA

Título original de la obra:
Ethics and the Limits of Philosophy

1.ª edición, 2016

Ilustración de cubierta: Auguste Rodin, *El pensador* (1880-1900)

Reservados todos los derechos. El contenido de esta obra está protegido por la Ley, que establece penas de prisión y/o multas, además de las correspondientes indemnizaciones por daños y perjuicios, para quienes reprodujeren, plagiaren, distribuyeren o comunicaren públicamente, en todo o en parte, una obra literaria, artística o científica, o su transformación, interpretación o ejecución artística fijada en cualquier tipo de soporte o comunicada a través de cualquier medio, sin la preceptiva autorización.

© All rights reserved. Authorised translation from the English language edition published by Routledge, a member of the Taylor & Francis Group
© Ediciones Cátedra (Grupo Anaya, S. A.), 2016
Juan Ignacio Luca de Tena, 15. 28027 Madrid
Depósito legal: M. 8.093-2016
I.S.B.N.: 978-84-376-3561-3
Printed in Spain

Índice

Prólogo a la edición de Routledge Classics. Jonathan Lear 9

Prefacio 19

Capítulo 1. La pregunta de Sócrates 23
Capítulo 2. El punto arquimédico 47
Capítulo 3. Los fundamentos: el bienestar 57
Capítulo 4. Los fundamentos: la razón práctica 83
Capítulo 5. Estilos de teoría ética 103
Capítulo 6. Teoría y prejuicio 127
Capítulo 7. El giro lingüístico 157
Capítulo 8. Conocimiento, ciencia y convergencia 169
Capítulo 9. Relativismo y reflexión 195
Capítulo 10. La peculiar institución de la moral 215

Epílogo 241

Comentario al texto. A. W. Moore 247

Prólogo a la edición de Routledge Classics

La ética y los límites de la filosofía nos propone que cambiemos la manera en la que hacemos ética. Bernard Williams sostiene que los filósofos no deberían seguir dando apoyo a lo que él llama «el sistema de la moral» —una estructura punitiva que pivota sobre la obligación, la censura y la culpa—, que se sustenta en la construcción de teorías morales, de la obligación, del castigo justo, etc., cada vez más complejas. Deberíamos centrarnos, en cambio, en esta otra cuestión: en la de en qué consistiría vivir nuestras vidas éticas con confianza, de manera legítima. En la generación posterior a la publicación de este libro se ha mantenido un debate vivo, a veces magistral, que ha cuestionado la concepción de la objetividad ética de Williams, así como su defensa del relativismo ético, su intento de dividir los conceptos éticos en densos y ligeros y si necesitaba categorías morales como las de lo correcto, la obligación y la justicia. En resumen, se ha criticado a Williams por casi todas sus tesis más importantes. Pero existe un acuerdo casi universal con respecto a que *La ética y los límites de la filosofía* ha marcado los temas de debate. No se puede entrar en el mundo del pensamiento ético contemporáneo sin tomar posición con respecto a las cuestiones planteadas por Williams. El libro también ha tenido un efecto perturbador en el estudio académico de la ética, en tanto que reta a profesores y estudiantes a que reflexionen sobre el tipo de actividad en que están inmersos y cuestionen su significado. Creo honestamente que se puede considerar que este libro es un clásico: para entender el devenir del

pensamiento ético en el siglo xx, hay que leer *La ética y los límites de la filosofía.*

A principios de los años ochenta, en Cambridge, tuve el privilegio de impartir una serie de seminarios sobre ética junto a Bernard Williams. En ellos, él puso a prueba las ideas que más tarde aparecerían en *La ética y los límites de la filosofía;* mi trabajo era plantearle retos cuando me era posible. Con la ventaja que da la mirada retrospectiva, desearía haber criticado con más vigor su defensa del *relativismo de la distancia,* con el cual Williams trató de articular una versión del relativismo que no sucumbiera a las objeciones obvias y que, según él pensaba, era el válido para la ética. «Una *confrontación real* entre dos perspectivas divergentes», estipula, «se da cuando, en un momento dado, hay un grupo de personas para quienes cada una de las perspectivas es una opción real» (págs. 199-200). Una opción real para un grupo es una perspectiva en la que ya se hallan o a la que podrían acceder en sus circunstancias históricas concretas, manteniendo su contacto con la realidad, evitando el autoengaño, etc. Por el contrario, una *confrontación* es *nocional* cuando un grupo puede ser consciente de las perspectivas alternativas, pero al menos una de esas perspectivas no supone una opción real. Como ejemplos de confrontación nocional, Williams cita la vida de un cacique de la Edad de Bronce o de un samurái de la Edad Media. La reflexión relativa a esos sistemas de vida puede ser valiosa, al «inspirar ciertos pensamientos relevantes para la vida moderna», pero «no es posible adoptar esas perspectivas» (pág. 200). Una perspectiva relativista, dice Williams, consiste en el punto de vista de que «solo a las confrontaciones reales puede aplicarse el lenguaje evaluativo: bueno, malo, correcto, incorrecto, etc.». Para Williams, no nos es legítimo evaluar moralmente a ese samurái, porque su modo de vida no es una opción real para nosotros. Yo he llegado a la conclusión de que esto no puede ser correcto.

Un problema es el de que las confrontaciones reales no pueden definirse en términos de opciones reales. No se puede, por ejemplo, comprender la brutalidad con que la cultura dominante estadounidense (así como los inmigrantes europeos) invadieron y destruyeron los modos de vida de los pueblos nativos, a no ser que uno comprenda que, para la cultura dominante, *no era posible* que la perspectiva alternativa fuera una opción real para ellos. La brutalidad de la confrontación se explica en parte porque, para cada bando, la perspectiva alternativa podía ser únicamente nocional. Una de las ironías crueles de aquella

situación fue que para Toro Sentado habría sido una opción más real adoptar la vida de un samurái medieval que la vida del general George Custer; mientras que, con respecto a Custer, habría sido tan poco realista pedirle que adoptara la perspectiva de Toro Sentado como que adoptara la de un samurái medieval. No se trata de que estas formas diversas de distancia tornen irrelevante la evaluación moral; son parte de la explicación causal de la atrocidad.

El caso que plantea un problema a la defensa de Williams del relativismo ético es el de una *confrontación real* (en el sentido de un conflicto en la vida real) entre partes para las que la perspectiva opuesta puede ser solo una opción *nocional*. Según las definiciones de Williams, estos casos no podrían existir. Sus intentos de encajar en su planteamiento este caso recalcitrante no me parece que funcionen. Él admite que podría haber conflictos en los que exista una asimetría: «Algunas versiones de la vida tecnológica moderna se han convertido en una opción real para miembros de sociedades tradicionales que sobreviven actualmente, pero su vida no es una opción real para nosotros, a pesar del entusiasmo nostálgico de algunos» (págs. 200-201).

Me preocupa que esta conceptualización refleje la perspectiva del vencedor. Después de todo, *¿quién dice* que la vida tecnológica contemporánea se ha convertido en una opción real para los supervivientes de una cultura tribal? Si pasáramos tiempo en la reserva de una tribu nativa americana hallaríamos que la cultura dominante sigue insistiendo en que presenta una opción real para la tribu, mientras que los miembros de la tribu una y otra vez se acaban dando de bruces con la imposibilidad de adoptar esa vida. Hace ahora ya más de un siglo que las tribus fueron confinadas a las reservas, han pasado generaciones, y aún hoy las dificultades que experimentan los pueblos nativos en relación con la cultura dominante son endémicas a la vida en las reservas. Esto resulta patético porque, contrariamente a lo que parece, no hay ninguna asimetría: ninguna perspectiva es una opción real para los otros, y eso no hace que el conflicto sea menos real y doloroso.

Williams dice, además, que *cualquier* conflicto contemporáneo tiene que ser un conflicto real, pero eso significa que o bien Williams pierde de vista su propia definición de la confrontación real (en términos de opciones reales) o bien tiene que insistir en que, precisamente porque el conflicto es contemporáneo, las opciones alternativas han de ser reales para quienes tienen la mala suerte de tener que afrontarlas.

> Hoy en día todas las confrontaciones entre culturas tienen que ser confrontaciones reales, y la existencia de sociedades tradicionales exóticas plantea cuestiones bastante distintas y difíciles de resolver, tal como la de si el resto del mundo puede o habría de usar el poder para preservarlas, como si fuesen especies en peligro de extinción; los antropólogos y otros investigadores de campo se encuentran a sí mismos ejerciendo el papel de guardas del coto (pág. 202).

Me preocupa que el relativismo de la distancia que Williams defiende nos incite a pensar que este es un problema fundamentalmente estético, circunscrito a nuestra altura de miras cultural, cuando se trata más bien, creo yo, de una obligación moral compleja y apremiante. Si este choque es realmente una colisión entre perspectivas alternativas que, sin lugar a dudas, *no* son opciones reales para ambas partes, entonces simplemente con llamarlas reales (porque son contemporáneas) no resolvemos el problema. Estamos ante un choque real de opciones nocionales. Y la pregunta es si el relativismo de la distancia fomenta que nos desentendamos de nuestras responsabilidades morales. Después de todo, ¿qué diferencia moral debería establecer la temporalidad? ¿Y si pudiéramos encontrarnos con un grupo nativo contemporáneo cuyo *ethos* fuera como el del samurái medieval? Este ejemplo no me parece caprichoso. Si es *absurdo* evaluar moralmente la manera de ser del samurái medieval, ¿en qué va a apoyarse la evaluación del samurái contemporáneo? La respuesta no puede ser que uno es un problema contemporáneo con el que tenemos que lidiar, mientras que el otro no lo es. Si el samurái contemporáneo nos plantea un problema moral, entonces parece que sería de gran valor poder evaluar el modo de vida del samurái medieval. Podría ayudarnos a encauzar los conflictos reales que surgen entre nosotros y aquellos con quienes solo podemos mantener una confrontación nocional.

Espero haber dicho lo suficiente para dejar claro —aunque sea mediante un único ejemplo— lo estimulante que resulta *La ética y los límites de la filosofía*. Estar en compañía de Bernard Williams es estar en compañía de un filósofo.

JONATHAN LEAR
Universidad de Chicago

Nota del traductor

A lo largo de todo el texto se ha seguido la convención de traducir *should* por «haber de» y *ought to* por «deber», pues la oposición entre ambos verbos es filosóficamente muy relevante para la posición del autor, y no podía quedar oscurecida. Sin duda, la pregunta socrática «¿Cómo ha de vivir uno?» desempeña un papel central en la argumentación del libro, y es fundamental entender que no puede construirse con el verbo «deber», en el sentido de obligación propio del sistema de la moral; mientras que la perífrasis «haber de» puede incorporar un matiz de conveniencia e intención futura que la hace más apta para ese uso.

Quiero agradecer la inestimable colaboración en esta traducción de Miracle Garrido, cuya minuciosidad ha salvado al texto de determinados errores y ha mejorado mucho su legibilidad y estilo. A ella se debe también la confrontación de la presente versión con la traducción francesa de Marie-Anne Lescourret, publicada por Éditions Gallimard con un prólogo del mismo Williams.

S. R.

Para Jacob

How cold the vacancy
When the phantoms are gone and the shaken realist
First sees reality. The mortal no
Has its emptiness and tragic expirations.
The tragedy, however, may have begun,
Again, in the imagination's new beginning,
In the yes of the realist spoken because he must
Say yes, spoken because under every no
Lay a passion for yes that had never been broken.

[Qué frío el vacío
cuando los fantasmas se han ido y el estremecido realista
ve por primera vez la realidad. El mortal no
tiene su vacío y sus vencimientos trágicos.
La tragedia, sin embargo, puede haber empezado,
una vez más, en el nuevo comienzo de la imaginación,
en el sí que el realista expresa porque debe
decir que sí, porque por debajo de todo no
subyace una pasión por el sí que nunca se haya roto].

WALLACE STEVENS, «Esthétique du Mal»

Quand on n'a pas de caractère, il faut bien se donner une méthode.

[Cuando uno carece de carácter, bien hace en darse un método].

ALBERT CAMUS, *La Chute*

Prefacio

Este libro se ocupa, principalmente, de cuál es el estado actual de la filosofía moral, y no de cuál podría ser; por lo que, dado que no creo que este estado sea el que debiera, buena parte del libro será crítica con la filosofía contemporánea. El libro plantea, además, la cuestión de en qué medida puede la filosofía ayudarnos a recrear una vida ética; y trataré de mostrar que puede, al menos, ayudarnos a comprenderla. Al constatar, y deplorar, el estado de cosas vigente, espero introducir una imagen del pensamiento ético, y de un conjunto de ideas que se le aplican, que puedan ayudarnos a pensar cómo podría ser.

Me gustaría hacer dos breves observaciones preliminares. Puede parecer sorprendente que un estudio de la filosofía moral contemporánea se ocupe, especialmente en los tres primeros capítulos, de examinar ciertas ideas procedentes del pensamiento griego antiguo. No es solo por devoción de la filosofía hacia su propia historia. Tengo razones específicas para hacerlo, que se pondrán de manifiesto, espero, según avance el libro (trato de aclararlo en el Epílogo). Ciertamente, la idea no es que las exigencias del mundo moderno al pensamiento ético no se diferencien en nada de las del mundo antiguo. Por el contrario, mi conclusión es que las exigencias del mundo moderno al pensamiento ético no tienen precedentes, y que las ideas sobre la racionalidad que incorpora la mayor parte de la filosofía moral contemporánea no pueden satisfacerlas; pero también creo que una actualización del pensamiento antiguo, muy modificado, sí podría conseguirlo.

La segunda observación es sobre el estilo. La filosofía de este libro no cabe duda de que pude calificarse, en un sentido amplio, de «analítica», como lo es gran parte de la filosofía reciente que discute. En realidad, considero que se trata de una cuestión de estilo, y la limitación que establece procede solo del hecho de que el estilo tiene que dejar su marca en la materia tratada. No hay un objeto distintivo del que se ocupe la filosofía moral analítica, si se compara con otros tipos de filosofía moral. Lo que distingue a la filosofía analítica de otros tipos de filosofía contemporánea (aunque no de gran parte de la filosofía de otras épocas) es un determinado modo de proceder, que implica la argumentación, el establecimiento de distinciones y, en la medida en que se acuerde de intentar alcanzarlo y lo consiga, un lenguaje moderadamente llano. Como alternativa a este lenguaje llano, la filosofía analítica distingue nítidamente entre oscuridad y tecnicismo: rechaza siempre la oscuridad, pero a veces encuentra necesario el tecnicismo. Este rasgo enfurece particularmente a algunos de sus enemigos, quienes, queriendo que la filosofía sea a la vez profunda y accesible, se molestan ante el tecnicismo pero se reconfortan en la oscuridad.

El objetivo de la filosofía analítica, como afirma siempre ella misma, es ser clara. Yo no estoy del todo seguro de que pueda reivindicar la titularidad de esta pretensión, y menos aún que sea de su exclusiva propiedad. Pero no quiero debatir esto aquí, en parte porque hacerlo supone dejar de debatir otras cuestiones, y también porque no me preocupa mucho qué consideración reciba el libro con respecto a si es o no filosofía analítica; me limito a reconocer que será considerado como tal. Lo que sí me importa es que sea lo que yo llamo «claro». Propongo a lo largo del libro que determinadas interpretaciones de la razón y de la comprensión clara como racionalidad discursiva han perjudicado al pensamiento ético y han distorsionado nuestra concepción de este. Pero, para que afirmaciones de este tipo resulten convincentes cuando vienen de un filósofo, es mejor exponerlas con cierto grado de racionalidad discursiva y siguiendo un orden argumentativo, y esto es lo que he tratado de hacer. No dudo de que no siempre lo habré conseguido, y de que hay muchas cosas que son oscuras, aunque he intentado esclarecerlas. Puedo reconocer esto con más convencimiento que el hecho de que algunas cosas son oscuras debido a que he intentado clarificarlas de esta manera, pero, sin duda, esto también es cierto.

Estoy en deuda con muchas personas por la ayuda que me han prestado, si bien ninguna de ellas es responsable de los resultados. Pude poner a prueba diversas versiones preliminares de mis críticas a la teoría ética en determinadas conferencias que tuve el honor de ser invitado a impartir: las Tanner Lectures en el Brasenose College de Oxford, las Thalheimer Lectures en la Universidad John Hopkins y las Gregynog Lectures en la Universidad de Gales en Aberystwyth. Agradezco a todos los presentes sus comentarios y críticas. Tuve la oportunidad de impartir un seminario sobre filosofía moral en la Universidad de Princeton en 1978, como Senior Visiting Fellow in the Humanities, y me fue provechoso el debate con muchas personas, en particular con Thomas Nagel y Tim Scanlon. Ronald Dworkin es desde hace tiempo un crítico amigable, minucioso y siempre insatisfecho. Borradores de alguna parte o de todo este libro han sido leídos por Geoffrey Hawthorn, Derek Parfit, Jonathan Lear y Amartya Sen, con todos los cuales estoy en deuda por sus comentarios. Agradezco la asistencia técnica de Mark Sacks y Peter Burbidge.

La cita de Wallace Stevens procede del libro *The Collected Poems,* 1954, reimpreso aquí con permiso de sus editores, Alfred A. Knopf, Inc. (Nueva York) y Faber and Faber (Londres).

Cambridge, Inglaterra

Capítulo 1

La pregunta de Sócrates

No estamos ante una cuestión trivial, decía Sócrates, pues de lo que hablamos es de cómo ha de vivir uno. O así lo cuenta Platón en uno de sus primeros libros dedicados a este tema[1]. Platón pensaba que la filosofía podía responder a esta pregunta. Como Sócrates, tenía esperanzas en que uno pudiera orientar su vida, y si fuese necesario reorientarla, por medio de una comprensión distintivamente filosófica; es decir, general y abstracta, racional y reflexiva, interesada en lo que puede conocerse por medio de distintos tipos de investigación.

Los objetivos de la filosofía moral, así como toda esperanza que esta pueda tener en ser digna de suscitar un interés serio, van unidos al destino de la pregunta de Sócrates, incluso aunque no sea razonable esperar que, por sí misma, la filosofía pueda responderla. Con respecto a esta esperanza, hay dos cosas que cabe mencionar de manera preliminar. Una, en concreto, ha de tenerla presente el autor: ¿qué alcance tiene su afirmación de que un tipo particular de escritura abstracta, argumentativa, ha de ser digna de atención seria cuando están en cuestión estas grandes preguntas? Hay otros libros que se ocupan de este asunto; si vamos a ello, casi todos los que tienen algo de bueno y que se preocupan por la vida humana. Esta es una apreciación relevante para

[1] Sócrates, *República* 352d.

quien escribe sobre temas filosóficos, aunque no crea que su relación con la pregunta de Sócrates consista en tratar de contestarla.

La otra observación inicial va dirigida al lector. Sería algo muy importante que la filosofía pudiera dar respuesta a la pregunta. ¿Cómo puede ser que una *disciplina* que se estudia en las universidades (pero no solo en ellas), que posee un extenso corpus bibliográfico de carácter técnico, pueda darnos lo que podríamos reconocer como una respuesta a las cuestiones fundamentales de la vida? Es difícil concebir cómo podría ser esto así, a menos que, como creía Sócrates, el lector pudiera reconocer la respuesta como una que él podría haberse dado a sí mismo. Pero ¿cómo puede conseguirlo? ¿Y qué relación tendría esto con la existencia de la disciplina? Para Sócrates, no existía esta disciplina; él simplemente hablaba con sus amigos de modo llano, y los escritores a los que hacía mención (al menos con cierto respeto) eran poetas. Pero en tan solo una generación Platón había unido el estudio de la filosofía moral a las difíciles disciplinas matemáticas, y dos generaciones después ya había tratados sobre la materia; en concreto, la *Ética* de Aristóteles, que sigue siendo uno de los más iluminadores[2].

A algunos filósofos les gustaría poder regresar a la situación de Sócrates y volver a empezar; cuestionarse de un modo reflexivo el sentido común y nuestros intereses morales o éticos, sin la carga de los textos y de la tradición filosófica. Hay elementos a favor de esta posición, y en este libro trataré de seguir su rastro hasta el punto de llevar a cabo una investigación en la que espero implicar al lector. En otro orden de ideas, sin embargo, carece de fundamento suponer que uno puede o debiera intentar separarse de las prácticas de la disciplina. Lo que proporciona a una investigación su carácter filosófico es la generalidad reflexiva y un estilo argumentativo que pretende ser racionalmente persuasivo. Sería insensato olvidar que muchas personas perspicaces y reflexivas se han ocupado ya de formular y debatir estas cuestiones. La filosofía moral se centra en unos problemas y no en otros a causa de su historia y de sus prácticas actuales. Además, no es baladí que exista una tradición de actividad, en parte técnica, en otras subdisciplinas filosóficas, tales como la lógica, la teoría del significado y la filosofía de la mente. Aunque, más allá de la lógica matemática muy pocas producen «resultados», hay mucho que aprender sobre el

---

[2] Sobre la obra de Aristóteles, véase la nota 6 del capítulo 3.

estado de la disciplina, y parte de ello incide significativamente en la filosofía moral.

Hay otra razón para no perder de vista que vivimos ahora y no en la época y las circunstancias de Sócrates. Para él y para Platón, existía un rasgo especial en la filosofía: era una disciplina reflexiva que se mantenía separada de las prácticas y los argumentos cotidianos para definir y criticar las actitudes implicadas por ellos. Pero la vida moderna es tan generalizadamente reflexiva, y es tan fundamental para sus instituciones tener un alto grado de autoconsciencia, que esas cualidades no pueden constituir la característica distintiva de la filosofía: lo que la distinga, por ejemplo, del derecho, que cada vez es más consciente de ser una creación social; o de la medicina, a la que no le queda más remedio que concebirse simultáneamente como cuidado, negocio y ciencia aplicada; por no hablar de la ficción, que incluso en sus formas más populares necesita ser consciente de su carácter ficticio. En el mundo moderno, la filosofía no puede presentarse como la disciplina de la reflexividad, aunque puede ser capaz de hacer un uso especial de esta.

Este libro intentará dar idea de los desarrollos más importantes acaecidos en la filosofía moral, pero lo hará indagando en sus problemas, en aquellas direcciones que me parecen más interesantes. Espero ser fiel a lo que afirman las posiciones que expongo, pero seré decididamente selectivo. La cuestión no es solo que mi concepción de la disciplina sea diferente de la de cualquier otro (cosa que parece que debe ser así para que el libro merezca ser leído), ni lo representativa que sea, sino, más bien, que no me preocuparé especialmente por esto. Hay un aspecto, al menos, en el que este libro no es representativo de las maneras principales en las que hoy en día se aborda la materia, por lo menos en el mundo anglófono: es más escéptico que gran parte de esa filosofía con respecto al poder de la filosofía misma, y también es más escéptico sobre la moral.

Que la filosofía moral tenga unos u otros objetivos depende de sus propios resultados. Dado que sus investigaciones son, de hecho, reflexivas y generales, y están interesadas en lo que puede conocerse, tienen que intentar dar una caracterización de qué ha de formar parte de una respuesta a la pregunta de Sócrates, la parte que correspondería al conocimiento propio de las ciencias; hasta dónde puede llevarnos la indagación puramente racional; las diferencias que podríamos esperar en las respuestas si la pregunta se hiciera en una sociedad o en otra; el margen que, después de todo ello, tiene que dejarse a la decisión personal. La

reflexión filosófica tiene que considerar, pues, lo que conlleva responder a estas o a otras preguntas prácticas menos generales, y preguntarse qué poderes de la mente y qué formas de conocimiento podría requerir. Y también tendrá que considerarse, en este proceso, el lugar de la filosofía misma.

Podría parecer que estamos atrapados en un círculo: la filosofía, al preguntarse cómo se podría contestar a la pregunta de Sócrates, determina su propio lugar al dar la respuesta. Pero esto no es un círculo, sino un avance. La filosofía parte de cuestiones por las que, desde cualquier punto de vista, puede y debe preguntarse, sobre las opciones que tenemos para hallar la mejor manera de vivir. En el camino descubre lo mucho que puede contribuir mediante el método discursivo del análisis y la argumentación, la insatisfacción crítica y la comparación imaginativa entre posibilidades, lo que constituye su forma característica de intentar enriquecer nuestras fuentes cotidianas de conocimiento histórico y personal.

La pregunta de Sócrates es el mejor punto de partida para la filosofía moral. Es mejor que «¿Cuál es nuestro deber?» o «¿Cómo podemos ser buenos?», o incluso que «¿Cómo podemos ser felices?». Cada una de estas preguntas presupone muchas cosas, aunque no todo el mundo esté de acuerdo en qué es lo que está implícito. En el caso de la última pregunta, algunas personas, como es el caso de quienes quieren empezar con la primera pregunta, pensarán que se parte del lugar equivocado, ignorando las cuestiones distintivas de la moral; otros pueden encontrarla más bien optimista. La pregunta de Sócrates es neutral a este respecto y en muchos otros. La primera cosa que hemos de hacer es preguntarnos qué implicaciones tiene la pregunta de Sócrates y cuánto presuponemos con tan siquiera asumir que puede ser útil plantearla.

«¿Cómo ha de vivir uno?». La generalidad de este *uno* ya supone una declaración. La lengua griega ni siquiera utiliza el pronombre indefinido *uno:* la fórmula es impersonal*. La implicación es que puede decirse algo relevante o útil para cualquiera, en general, y esto implica que puede decirse algo general, algo que abarca o conforma las ambiciones individuales con las que cada cual puede abordar la pregunta de

* También en la mayoría de traducciones directas al castellano. *[N. del T.]*.

«¿Cómo he de vivir?». (Una implicación más amplia que puede hallarse fácilmente en esta generalidad es que la pregunta nos conduce, de modo natural, fuera de los intereses del ego. Volveremos sobre esto más adelante). Esta es una manera en que la pregunta de Sócrates va más allá del cotidiano «¿Qué voy a hacer?». Otra implicación es que no es inmediata; no se refiere a lo que he de hacer ahora, o a continuación, sino a un modo de vida. Los propios griegos tenían en alta estima la idea de que, en consecuencia, esta pregunta tenía que referirse al conjunto de la vida y que una buena manera de vivir tenía que partir de lo que, cuando llegase a su fin, se vería como una vida buena. Impresionados por el poder de la fortuna para arruinar lo que parecía la vida mejor conformada, algunos de ellos, y Sócrates el primero, buscaron un diseño racional de la vida que redujera el poder de la fortuna y estuviera, lo máximo posible, a salvo de la suerte[3]. Este también ha sido, en diferentes formas, el objetivo del pensamiento posterior. La idea de que uno tiene que pensar, en este nivel muy general, en *toda una vida* puede parecer menos apremiante para algunos de nosotros de lo que lo era para Sócrates. Pero esta pregunta aún nos exige reflexionar sobre nuestra *vida como un todo,* a partir de cualquier aspecto y en todos los niveles, incluso si no ponemos tanto énfasis como los griegos en cómo puede terminar.

La frase impersonal griega, traducida como *uno ha de* no solo guarda silencio sobre la persona cuya vida está en cuestión, sino que tampoco comporta compromiso alguno con qué tipo de consideraciones se aplican a la pregunta, lo cual resulta muy fructífero. «¿Cómo he de vivir?» no significa «¿qué vida debo vivir desde un punto de vista moral?». Esta es la razón por la que la pregunta de Sócrates es un punto de partida diferente del resto de preguntas que he mencionado sobre el deber o sobre una vida en la que uno sea bueno. Puede ser lo mismo en tanto que pregunta sobre la vida buena, una vida que merezca ser vivida, pero esa noción en sí misma no introduce ninguna exigencia distintivamente moral. Puede resultar que, como Sócrates pensaba y la mayoría de nosotros todavía esperamos, una vida buena sea también la vida de una

---

[3] Martha C. Nussbaum, *The Fragility of Goodness* (Nueva York, Cambridge University Press, 1985) [*La fragilidad del bien,* trad. de A. Ballesteros (Madrid, Visor, 1995; A. Machado Libros, 2004)], explora esta y otras nociones afines de la literatura y la filosofía antiguas. Para la concepción según la cual la moral trasciende la suerte, véase el capítulo 10.

buena persona *(tiene que serlo* es lo que Sócrates creía; *puede serlo* es lo que muchos de nosotros esperamos). Pero, si es así o no, se verá posteriormente. *Ha de* es simplemente *ha de,* y en sí no es diferente, en esta versión general de la pregunta, de lo que es en cualquier pregunta banal, como «¿Qué he de hacer ahora?».

Algunos filósofos han supuesto que no podemos comenzar por este tipo general e indeterminado de pregunta práctica, porque preguntas como «¿Qué he de hacer?», «¿Cuál es para mí el mejor modo de vivir?», etc., son *ambiguas* y admiten una lectura tanto moral como no moral. Desde esta posición, lo primero que uno tiene que hacer es decidir a cuál de los dos aspectos se refiere la pregunta, y hasta ese momento uno no podría ni siquiera empezar a responderla. Esto es un error. El análisis de los significados no requiere las categorías semánticas de «moral» o «no moral». Por supuesto, si alguien dice de otro «es un hombre bueno», cabe preguntarse si el hablante quiere decir que es moralmente bueno, en contraste, por ejemplo, con que sea un hombre bueno para una misión de combate. Pero el hecho de que uno pueda dar esas interpretaciones diversas no implica más un sentido moral de «bueno» o de «hombre bueno» que un sentido militar (o futbolístico, etc.).

Por supuesto, uno puede preguntarse, en una ocasión dada: «¿Qué he de hacer desde un punto de vista ético?» o «¿Qué he de hacer desde el punto de vista del interés propio?». Son cuestiones que conciernen a los resultados de deliberaciones previas subyacentes, y nos invitan a revisar un tipo particular de consideración entre todas las que guardan relación con la pregunta y a pensar en las consecuencias de ese tipo de planteamientos. Del mismo modo, también puedo preguntarme qué he de hacer teniendo en cuenta solo consideraciones económicas, políticas o familiares. Al final de todo esto, surge la cuestión de «¿Qué he de hacer, en definitiva?». Hay solo un tipo de pregunta que plantearse sobre qué hacer, de la que la pregunta de Sócrates es un ejemplo muy general, y las consideraciones morales son solo uno de los tipos de consideraciones relevantes para responderla[4].

Ya me he referido a las consideraciones «morales», usando esta palabra de una manera general, que se corresponde con uno de los nom-

---

[4] Si algunos filósofos han hallado dificultades en esta caracterización obvia, puede ser porque presuponen que esta pregunta «global» habría de contestarse apelando a un solo tipo de consideración. Defenderé más adelante, en el presente capítulo, que este presupuesto es erróneo.

bres que se le dan invariablemente a la disciplina: filosofía moral. Pero esta tiene otro nombre, «ética», término al que corresponde la noción de consideración ética. Por origen, la diferencia entre ambos términos es que una palabra procede del latín y la otra del griego, y ambas se relacionan con una palabra que significa *disposición* o *costumbre*. Una diferencia es que el término latino del que procede «moral» enfatiza más el sentido de expectativa social, mientras que el griego favorece el de carácter individual. Pero la palabra «moral» ha adquirido hoy en día un contenido más distintivo, y voy a proponer que la moral debería entenderse como un desarrollo particular de lo ético, que tiene un significado especial en la cultura occidental moderna y enfatiza de una manera peculiar ciertas nociones éticas más que otras; desarrolla, en particular, una noción especial de obligación, e incorpora ciertos presupuestos peculiares. A la vista de estas características, es también algo que, a mi entender, deberíamos tratar con un especial escepticismo. Por lo tanto, de ahora en adelante usaré habitualmente «ético» como el término de significado más amplio que designa aquello sobre lo que versa esta disciplina, y «moral», como adjetivo y como sustantivo, para el sistema más concreto, a cuyas peculiaridades nos referiremos más adelante.

En vez de definir qué cuenta exactamente como una consideración ética, intentaré elaborar el alcance de la noción de lo ético. No importa que la noción sea vaga. En realidad, es la *moral,* el sistema especial, la que requiere unos límites nítidos (al exigir sentidos «morales» y «no morales» de las palabras, por ejemplo). Esto deriva de sus presupuestos específicos. Sin ellos, podemos admitir la existencia de toda una serie de consideraciones que caen bajo la noción de lo ético, así como comprender por qué esta esfera no está nítidamente delimitada.

Dentro de su dominio se encuentra la noción de obligación. Hay un conjunto bastante variado de consideraciones que normalmente concebimos como obligaciones; me ocuparé más adelante (en el capítulo 10) de la cuestión de por qué esto ha de ser así. Un tipo que nos resulta familiar es el de la obligación que alguien puede imponerse a sí mismo al hacer una promesa. Está también la idea de deber. El uso más habitual de esta palabra hoy en día suele darse en contextos institucionales, en los que suele haber una serie de tareas o cometidos. Más allá de eso, es característico que los deberes dependan de una función, un cargo o un vínculo, tales como los que derivan de nuestra «posición», por usar el término que Bradley utilizó en el título de un

famoso artículo[5]. En el caso de deberes propios del ejercicio de una profesión, pueden ser fruto de la aceptación voluntaria de un empleo, pero en general los deberes, y la mayoría de las obligaciones que no derivan de promesas, no se contraen de forma voluntaria.

En el pensamiento de Kant y de otros pensadores influidos por él, todas las consideraciones genuinamente morales descansan en último término, y de manera profunda, en la voluntad del agente. No se me puede *requerir* solo por mi posición en una estructura social —por el hecho de ser hijo de cierta persona, por ejemplo— que actúe de una determinada manera, si ese *requerimiento* es de tipo moral, ni puede esta posición simplemente reflejar una compulsión psicológica o sanciones sociales o legales. Actuar moralmente es actuar autónomamente, no por presión social. Esto refleja algunas características propias del subsistema de la moral. Frente a ello, ha existido en toda sociedad, y la nuestra no es una excepción, la idea reconociblemente ética de que uno puede estar sometido a un requerimiento de ese tipo simplemente por ser quien es o por su situación social. Puede ser un tipo de consideración que pocas personas en las sociedades occidentales querrían aceptar hoy en día, pero casi todo el mundo la aceptaba en el pasado, y no hay ninguna necesidad en la exigencia de que todo requerimiento de este tipo, tras ser sometido al escrutinio racional, tenga que ser o abandonado o convertido en un compromiso voluntario. Tal exigencia, como otros rasgos distintivos de la moral, está estrechamente conectada a procesos de modernización: representa una comprensión en términos éticos del proceso que, en el mundo de las relaciones jurídicas, Maine llamó el paso del estatus al contrato. Se corresponde también con una cambiante concepción del yo que mantiene relaciones éticas[6].

La obligación y el deber nos hacen echar la mirada atrás, o por lo menos a los lados. Los actos que exigen, suponiendo que uno está deli-

---

[5] F. H. Bradley, «My Station and Its Duties», en *Ethical Studies*, 2.ª ed. (Indianápolis, Bobbs-Merril, 1951); publicado originalmente en 1876.

[6] La idea aparece en Alasdair MacIntyre, *After Virtue* (Notre Dame, University of Notre Dame Press, 1981) [*Tras la virtud*, trad. de A. Valcárcel (Barcelona, Crítica, 1987; Austral, 2013)], y en Michael Sandel, *Liberalism and the Limits of Justice* (Nueva York, Cambridge University Press, 1982) [*El liberalismo y los límites de la justicia*, trad. de M. L. Melón, J. F. Malem Seña y E. Garzón Valdés (Barcelona, Gedisa, 2000)]. El énfasis de Sandel en un «yo socialmente constituido» se enfrenta a dificultades que conocen bien los neohegelianos. Véase también el capítulo 10, nota 16 y, sobre MacIntyre, la nota 13 de este capítulo.

berando sobre qué hacer, residen en el futuro, pero las razones favorables a estos actos se hallan en la promesa que he realizado, el empleo que he aceptado, la posición que ocupo. Otro tipo de consideración ética nos hace mirar hacia delante, a los resultados de los actos que puedo realizar. «Será lo mejor» puede tomarse como la forma general de este tipo de consideración. Según una manera de entender esto, especialmente importante en la teoría filosófica, *lo mejor* se mide por el grado en que las personas consiguen lo que quieren, se tornan más felices, o alguna consideración similar. Esta es el área de la ética del bienestar o del utilitarismo (examinaré estas concepciones en los capítulos 5 y 6). Pero esa es solo una versión. G. E. Moore también pensó que el tipo prospectivo de consideración era fundamental, pero aceptó que otros conceptos además de la satisfacción —tales como la amistad o la experiencia de la belleza— contaran entre las consecuencias buenas. A esto se debe que su teoría fuera tan atractiva para el grupo de Bloomsbury: permitía rechazar, a la vez, la estrechez del deber y la vulgaridad del utilitarismo.

Existe otro tipo de consideración ética que presenta una acción como éticamente relevante. Hay una amplia gama de características éticas que permiten elegir o rechazar determinadas acciones. Una acción particular puede rechazarse porque supondría un robo o un asesinato, por ejemplo, o sería engañosa o deshonrosa o, menos drásticamente, porque decepcionaría a alguien. Estas descripciones —y hay muchas— operan en diferentes niveles; así, una acción puede ser deshonrosa porque es engañosa.

Hay diversas virtudes, estrechamente relacionadas con esas descripciones, que permiten elegir o rechazar acciones. La virtud es una disposición del carácter a elegir o rechazar acciones en función de su cualidad ética. La palabra «virtud» ha adquirido, en gran medida, asociaciones cómicas e indeseables, y pocos la usan hoy en día, a excepción hecha de los filósofos, pero no hay otra palabra mejor y la filosofía moral tiene que usarla. Se puede esperar que, una vez se restablezca su significado propio, recuperará su respetabilidad. En su uso apropiado, una virtud es una disposición del carácter éticamente admirable, y cubre una clase amplia de características; pero como sucede a menudo en estas materias, los límites de esa clase no son nítidos y no necesitan serlo. Algunas características personales deseables, tales como el ser sexualmente atractivo, ciertamente no cuentan como virtudes. Esta puede ser una cuestión de personalidad (algunas personas tienen un carácter sexualmente atrac-

tivo), pero no tiene por qué ser una virtud y no está considerada como tal, al igual que ocurre con la posesión de un oído perfecto. Por otra parte, las virtudes son siempre más que meras destrezas, dado que implican patrones característicos de deseo y motivación. Se puede ser un buen pianista y no tener ningún deseo de tocar, pero, si uno es generoso o imparcial, esas cualidades ayudan a determinar, en los contextos apropiados, lo que querrá hacer.

Esto no significa que no pueda hacerse un mal uso de las virtudes, como en el caso de las llamadas virtudes ejecutivas —el coraje, por ejemplo, o el autocontrol—, que no implican la consecución de objetivos propios, sino que contribuyen a alcanzar otras metas. Se trata, no obstante, de virtudes, en tanto que son rasgos de carácter y no están conectadas con la persecución de unos objetivos, como lo está la mera posesión de una destreza. Para Sócrates, las virtudes no pueden usarse mal, y de hecho sostuvo incluso algo más categórico: que es imposible que las personas que poseen cierta virtud actúen peor que si no la tuvieran. Esto le llevó, consecuentemente, a creer que básicamente solo hay una virtud, el poder del buen juicio. No hace falta que estemos de acuerdo en esto. Y, lo que es más importante, tampoco hemos de hacerlo con respecto a lo que motiva esas ideas, que es la búsqueda de algo que en la vida de un individuo pueda ser bueno *sin matices,* bueno en todas las circunstancias posibles. Esa búsqueda tiene también sus expresiones modernas, y nos encontraremos con una de ellas al ocuparnos de las preocupaciones específicas de la moral.

La noción de virtud es una noción tradicional de la filosofía moral, que fue excluida de los debates durante un tiempo, hasta que recientemente diversos escritores han destacado con acierto su importancia[7]. Si

---

[7] P. T. Geach, *The Virtues: The Stanton Lectures, 1973-1974* (Nueva York, Cambridge University Press, 1977) [*Las virtudes,* trad. de J. V. Arregui y C. Rodríguez Ledesma (Pamplona, EUNSA, 1993)]; Philippa Foot, *Virtues and Vices* (Berkeley: University of California Press, 1978) [*Las virtudes y los vicios,* trad. de C. Martínez (México, UNAM-IIF, 1994)]; James D. Wallace, *Virtues and Vices* (Ithaca, Cornell University Press, 1978); MacIntyre, *Tras la virtud.* Las razones de su desatención se hallan en una concepción restringida de los intereses éticos y en la concentración en las preocupaciones de la moral; también puede ser que el estudio de las virtudes se haya asociado a presupuestos religiosos (que tienen una presencia muy marcada en la obra de Geach). Hay una objeción a la idea de virtud que cabe tomarse en serio, que carga las tintas en la invocación de la noción de carácter, una noción que para nosotros ya no tendría ningún sentido, o no el suficiente. Digo algo sobre esta cuestión en el Epílogo. Creo que la objeción, si se desarrolla

alguien tiene una virtud determinada, ello afecta a cómo delibera. Es necesario que seamos claros, sin embargo, sobre las maneras en las que esa virtud puede afectar a la deliberación. Un asunto importante es que el término que designa una virtud no suele aparecer en el contenido de la deliberación. Quien tiene una virtud particular realiza unas acciones porque caen bajo determinadas descripciones y evita otras porque caen bajo otras descripciones. Esa persona es descrita en términos de la virtud, como sucede con sus acciones: es una persona justa o valiente quien hace cosas justas o valientes. Pero, y he aquí la cuestión, casi nunca sucede que la descripción que se aplica al agente y a la acción sea la misma que aquella en términos de la cual el agente elige su acción. «Justo» es, sin duda, uno de los pocos casos en los que esto es así, y una persona justa o equitativa es alguien que elige unas acciones porque son justas y rechaza otras porque son injustas o poco equitativas. Pero, por regla general, una persona valiente no elige unos actos porque sean valientes, y es público y notorio que una persona modesta no actúa por mor de la modestia. La persona benevolente o compasiva hace cosas benevolentes, pero las hace bajo otras descripciones, tales como que otra persona «lo necesita», «la animará», «dejará de sufrir». La descripción de la virtud no figura en la descripción que aparece en la consideración. Además, normalmente no hay ningún concepto ético que caracterice las deliberaciones de una persona que tiene una virtud particular. Más bien, si un agente tiene una virtud particular, entonces determinados conjuntos de hechos se tornan consideraciones éticas para ese agente porque tiene esa virtud. El camino que va de las consideraciones éticas que pondera una persona virtuosa a la descripción de la virtud es muy tortuoso, y viene definido y marcado por el impacto de la autoconsciencia.

Ese mismo impacto, de hecho, puede haber contribuido a que las virtudes sean impopulares como concepción ética. El *cultivo* de las virtudes solía consistir, en buena medida, en debatir sobre ellas. En tercera persona, ese ejercicio, aunque no sea con este nombre, nos resulta muy familiar: supone una buena parte de la socialización o de la educación moral o de la educación sin más. Como ejercicio en primera persona, sin embargo, el cultivo de las virtudes tiene algo de sospechoso, de mojigatería o de autoengaño. No es solo que pensar de esta manera es

---

lo suficiente, constituye una objeción al pensamiento ético mismo más que a un modo determinado de elaborarlo.

pensar en uno mismo y no en el mundo o en los demás. Ciertas concepciones éticas, particularmente si son autocríticas, harán eso, sin duda. Más de un autor ha destacado recientemente la importancia de nuestra capacidad para tener deseos de segundo orden —deseos de tener ciertos deseos[8]—, así como su importancia para la reflexión ética y la consciencia práctica. La deliberación cuya meta es satisfacer esos deseos de segundo orden tiene que orientarse, en un grado especial, hacia el yo. El problema de cultivar las virtudes, si se ve como un ejercicio en primera persona y deliberativo, es más bien que el pensamiento no está lo suficientemente dirigido hacia el yo. Pensar en mis estados posibles en términos de las virtudes no es pensar en mis acciones, y no es distintivamente pensar en los términos en los que podría o habría de pensar sobre mis acciones: es, más bien, pensar en la manera en que otras personas podrían describir o comentar la manera en la que pienso sobre mis acciones y, si eso representa el contenido esencial de mis deliberaciones, parece sin duda que mi atención ética está mal encauzada. La lección de todo esto, sin embargo, no es que las virtudes no sean un concepto ético importante, sino que la importancia de un concepto ético no tiene por qué depender de que sea un elemento de la deliberación en primera persona. Las deliberaciones de las personas que son generosas o valientes, y también de las que tratan de ser más generosas y valientes, son diferentes de las de quienes no lo son, pero la diferencia no reside principalmente en que piensen sobre sí mismos en términos de generosidad o valentía.

Estas son, pues, algunas de las clases de conceptos y consideraciones de carácter ético. ¿Qué tipo de consideraciones se refieren a la acción *sin* ser consideraciones éticas? Hay un candidato obvio: las consideraciones egoístas, que se circunscriben al confort, el entusiasmo, la autoestima, el poder u otras ventajas para el agente. Es un lugar común

---

[8] Por ejemplo, Harry Frankfurt, «Freedom of the Will and the Concept of a Person», *Journal of Philosophy*, 67 (1971) [«Libertad de la voluntad y el concepto de persona», en *La importancia de lo que nos preocupa*, trad. de V. I. Weinstabl y S. M. de Hagen (Buenos Aires, Katz, 2006)]; Amartya Sen, «Choice, Orderings, and Morality», en Stephan Körner (ed.), *Practical Reason* (New Haven, Yale University Press, 1974); R. C. Jeffery, «Preference among Preferences», *Journal of Philosophy*, 71 (1974); A. O. Hirschman, *Shifting Involvements* (Princeton University Press, 1982), cap. 4.

contrastar este tipo de consideraciones con las éticas, sobre la base de ideas claramente razonables relativas a la finalidad de las prácticas éticas, la función que desempeñan en las sociedades humanas. Pero todavía hacen falta aquí algunas distinciones. Una es meramente verbal. Estamos interesados en la pregunta de Sócrates «¿Cómo ha de vivir uno?», y el egoísmo, en el sentido más básico y burdamente autointeresado, proporciona una respuesta inteligible, aunque la mayoría de nosotros la rechacemos. Es posible llamar «ético» a todo plan de vida que proporcione una respuesta inteligible a la pregunta de Sócrates. En este sentido, incluso el egoísmo más burdo sería una opción ética. Pero no veo necesario adoptar este uso. Por vaga que sea inicialmente, tenemos una concepción de lo ético que nos conecta de manera comprensible a nosotros y a nuestras acciones con las exigencias, las necesidades, las demandas, los deseos y, en general, las vidas de los demás, y es útil preservar esta concepción en lo que estamos dispuestos a denominar una consideración ética.

Sin embargo, el egoísmo puede dar un paso más y no quedarse en esta forma burda. Hay una teoría sobre cómo hemos de actuar que se conoce, de un modo un tanto confuso, como *egoísmo ético,* la cual afirma que toda persona debe buscar su propio interés. Difiere del egoísmo burdo en que es una posición reflexiva y adopta una concepción general sobre los intereses de las personas. No importa mucho si lo tenemos por un sistema ético, que es como se ve a sí mismo, o no. La pregunta importante es la de cómo contribuye a la idea de *consideración* ética. A primera vista, no parece que lo haga en absoluto, pues afirma que cada uno de nosotros debe actuar a partir de consideraciones no éticas. Si solo dice eso, es sencillamente dogmático: si las personas actúan, de hecho, por consideraciones que no son exclusivamente las que responden al autointerés, ¿por qué son irracionales al hacerlo? El efecto más probable de esta posición es que *deje abierto* el papel de las consideraciones éticas y se pregunte cuál es la relación de una vida que implica actuar por esas consideraciones con el autointerés.

Hay otra concepción, muy parecida a la anterior, pero que es muy diferente en realidad, que afirma también algo general: dice que *lo que debe pasar* es que cada cual busque su propio interés. Esta concepción es probable que tenga un efecto desestabilizador en las consideraciones que uno tiene en cuenta al actuar. Puede introducir una consideración que es ética en el sentido cotidiano. Si creo que lo que debe suceder es que las personas busquen su propio interés, entonces puedo tener

razones para promover ese estado de cosas, y esto puede llevar a que eche una mano a los demás al adoptar ese plan de acción. Este patrón de conducta puede entrar en conflicto con la mera búsqueda de mi propio interés.

De hecho, es bastante difícil mantener la mera creencia de que lo que debe suceder es que las personas busquen su propio interés. Es más natural apoyar esto en otra consideración: que *lo mejor* sería que todos actuáramos así. Esto puede expresarse diciendo que conductas tales como tratar de ser amable con los demás solo complica las cosas. Quien defiende esto (y se lo cree) ciertamente también acepta otras consideraciones éticas adicionales; por ejemplo, que es bueno que las personas consigan lo que quieren, y cree además que el mejor modo para que el mayor número de personas obtenga el máximo posible de lo que quieren es que cada cual busque lo que quiere. Esto es, por supuesto, lo que los defensores del capitalismo del *laissez-faire* solían sostener a principios del siglo XIX. Hay quien incluso lo mantiene a finales del siglo XX, a pesar del hecho evidente de que todos los sistemas económicos dependen de que las disposiciones de las personas que conforman una sociedad vayan más allá del autointerés. Quizá esta contradicción contribuye a explicar por qué algunos defensores del *laissez-faire* tienden a dar lecciones moralizantes, no solo a las personas que no buscan su propio interés, sino también a las que lo hacen.

Estamos contrastando consideraciones éticas y consideraciones egoístas. Pero, ¿no podría alguien querer la felicidad de otra persona? Por supuesto. Entonces, ¿no coincidiría el egoísmo, mi búsqueda de lo que quiero, con lo que se supone que es una consideración de tipo ético, la preocupación por la felicidad de otra persona? Sí, de nuevo; pero esto no resulta muy interesante a no ser que, en algún sentido más general y sistemático, las consideraciones egoístas y las éticas se combinen. Esta es una cuestión que abordaremos al tratar el problema de los fundamentos en el capítulo 3.

Teniendo en cuenta todo esto, se verá que la idea de lo ético, aunque sea vaga, tiene cierto contenido; no es una noción puramente formal. Una ilustración de ello la encontramos en un tipo diferente de consideración no ética que podríamos llamar lo *contraético*. Las motivaciones contraéticas, un fenómeno humano significativo, adoptan diversas formas, delimitadas por sus contrapartidas positivas en lo ético. La malevolencia, que es el motivo más conocido de este tipo, se asocia a menudo al placer del agente, que suele tenerse por su estado natural;

pero existe también una malevolencia pura y desinteresada, una malicia que trasciende incluso la necesidad del agente de estar cerca para disfrutar el daño que quiere infligir. Difiere de la contrajusticia, un deleite caprichoso producido por la injusticia, que es profundamente parasitario de su contrapartida ética, en el sentido de que hace falta primero una determinación cuidadosa de lo justo para orientarla. Con la malevolencia no sucede exactamente lo mismo. No es que la benevolencia venga primero para que la malevolencia tenga de donde partir, sino que ambas usan las mismas percepciones pero se apartan de ellas en direcciones diferentes. (Es por esto que, como señaló Nietzsche, la crueldad necesita compartir la sensibilidad de la simpatía, mientras que la brutalidad no lo precisa). Otras motivaciones contraéticas son parasitarias de la reputación o de la autoimagen emocional de lo ético más que de sus conclusiones. Como era de esperar, esto puede involucrar en especial a las virtudes. Que una acción sea cobarde no es algo que normalmente el agente encuentre como consideración a su favor, pero podría serlo y, en un sentido contraético, estaría al servicio de un masoquismo de la vergüenza.

Me he ocupado brevemente de las consideraciones egoístas y de las altruistas (guiadas por la benevolencia, por ejemplo, o por la equidad). Pero una cuestión de gran importancia para la ética es la de determinar hasta qué punto estas consideraciones pueden trascender el interés personal. ¿Contará como consideración ética si solo tenemos en cuenta los intereses y las necesidades de nuestra familia, de nuestra comunidad o de la nación? Ciertamente, estas lealtades locales han proporcionado la estructura de la vida de las personas y han delimitado el espacio público de la vida ética. Sin embargo, parece que únicamente un interés universal puede satisfacer ciertas exigencias éticas; un interés que se extienda a todos los seres humanos y que quizá vaya más allá de la raza humana. Este interés lo cultiva especialmente el subsistema de la moral, en la medida en que se suele pensar que ningún interés es verdaderamente moral a menos que tenga el cuño de la universalidad.

Para la moral, la circunscripción ética es siempre la misma: la circunscripción universal. La fidelidad a un grupo más pequeño, las lealtades a la familia o al país tendrían que justificarse de fuera hacia dentro, por medio de un argumento que explicara que es bueno que las personas tengan fidelidades que no son universales. (En los capítulos 5 y 6 abordaré los motivos y los peligros de este tipo de planteamiento, así como las diferentes caracterizaciones que se han propuesto de la

circunscripción universal). A un nivel más cotidiano (menos reflexivo, diría el crítico moral), la localización de lo ético puede desplazarse de un lado a otro de un contraste dado. En relación con mi interés personal, los intereses de la ciudad o de la nación pueden representar una exigencia ética, pero los intereses de la ciudad pueden considerarse autointeresados si la exigencia procede de cierta identificación más amplia. Esto se debe sencillamente a que las exigencias de la benevolencia o de la equidad siempre pueden plantear una demanda contra el autointerés; el *nosotros* puede representar al interés propio tanto como el *yo;* y quiénes somos *nosotros* depende del grado de identificación en un caso particular y de cómo se delimite el contraste.

He mencionado diversos tipos de consideraciones éticas y más de un tipo de consideración no ética. Tradicionalmente, la filosofía ha mostrado su deseo de reducir la diversidad interna a ambos extremos. Por un lado, ha tendido a concebir todas las consideraciones no éticas como reducibles al egoísmo, la forma más estrecha de autointerés. De hecho, ciertos filósofos han querido reducir esto a un tipo especial de interés egoísta, la búsqueda del placer. Kant, en particular, creía que toda acción que no responde a un principio moral se hace por placer. Este punto de vista se ha de distinguir de otra idea, según la cual todas las acciones, incluyendo las que se realizan por razones éticas, están igualmente motivadas por la búsqueda del placer. Es difícil que esta teoría, el hedonismo psicológico, no sea obviamente falsa o, si no, trivialmente vacua, como sucede si se limita a identificar la expectativa de placer del agente con cualquier cosa que este haga intencionalmente. Pero, en cualquier caso, esta teoría no hace ninguna contribución especial a la distinción entre lo ético y lo no ético. Si hubiera alguna versión verdadera e interesante del hedonismo psicológico, aquellas acciones que tienen motivaciones no éticas no formarían necesariamente ninguna clase especial de actividad orientada a la búsqueda del placer. Por otro lado, la posición de Kant contribuye a la distinción, al sostener que la acción moral es la única exenta del hedonismo psicológico; esta concepción es, sin duda, errónea[9]. Si escapamos a la influencia de esa teoría, podemos aceptar la verdad obvia de que hay diferentes tipos de motivación no ética —y, además, que

[9] La doctrina forma parte de la teoría de la libertad de Kant, la cual, si es posible hacerla inteligible, no sería nada fácil de salvar de la acusación de incoherencia. Para más comentarios sobre esto, véase el capítulo 4.

hay más de un tipo de motivación que actúa contra las consideraciones éticas—[10].

El deseo de reducir todas las consideraciones no éticas a un solo tipo es menos fuerte en la filosofía de hoy en día de lo que lo era en la época en que la filosofía moral se centraba fundamentalmente no tanto en las cuestiones de qué hay que hacer y cuál es la vida buena (las respuestas a tales preguntas pensaba que eran obvias), sino más bien en las motivaciones que conducían a la búsqueda de esos objetivos, contra las motivaciones del autointerés y el placer. El deseo de reducir todas las consideraciones *éticas* a un patrón es, por otro lado, tan fuerte como antaño, y diversas teorías intentan mostrar que uno u otro tipo de consideraciones éticas es básico y puede dar cuenta de los otros tipos. Hay quien toma como básica una noción de obligación o de deber, y el hecho de que, por ejemplo, tomemos como una consideración ética que determinado acto conduzca probablemente a las mejores consecuencias se explica en términos de que uno de nuestros deberes es producir las mejores consecuencias. Las teorías de este tipo se llaman «deontológicas». (Se dice a veces que este término procede de la palabra griega que equivale a deber, pero no hay ningún término en griego antiguo para referirse al deber; en realidad, procede de la palabra griega con la que se expresa lo que uno *tiene que* hacer).

A diferencia de las anteriores, otras teorías toman como primaria la idea de producir el mejor estado de cosas posible. A las teorías de este tipo se las conoce como «teleológicas». El ejemplo más importante es el que identifica la bondad de los resultados en términos de la felicidad de las personas o de que estas consigan lo que quieren o lo que prefieren. Esto, como ya he dicho, se denomina utilitarismo, aunque el término también se ha aplicado —como hizo, por ejemplo, Moore— a la noción más general de sistema teleológico[11]. Algunas de esas teorías reduccionistas solo nos dicen qué noción es racional, o más fiel a nuestra expe-

---

[10] He reflexionado sobre esto en *Morality: An Introduction to Ethics* (Nueva York, Harper and Row, 1972) [*Introducción a la ética,* trad. de M. Jiménez Redondo (Madrid, Cátedra, 1988)].

[11] La distinción entre deontológico y teleológico se introduce aquí de una forma muy indicativa. El interés de la distinción debe hallarse, probablemente, en un plano diferente, en un desacuerdo sobre cuál es la importancia de la moral; véase el capítulo 10. La distinción es solo una entre muchas de las que se han propuesto. Para una clasificación exhaustiva, véase W. K. Frankena, *Ethics,* 2.ª ed. (Englewood Cliffs, Prentice-Hall, 1973) [*Ética,* 1.ª ed., trad. de C. Gerhard (México, Ed. Hispano Americana, 1965)].

riencia ética, que tratemos como la noción fundamental. Otras teorías son más audaces y afirman que esas relaciones hay que descubrirlas en los significados de lo que decimos. Así, Moore sostuvo que «correcto» significa simplemente «lo que produce el mayor bien»[12]. La filosofía de Moore está afectada de una modesta cautela tal que satura su prosa con numerosas salvedades, las cuales, no obstante, casi nunca lo salvan de tremendos errores, además de que como análisis del significado de las palabras es simplemente falsa. En general, si las teorías de este tipo se presentan como descriptivas, como caracterizaciones de cosas que consideramos equivalentes, están todas igualmente equivocadas. Usamos una gran variedad de consideraciones éticas diferentes, que son genuinamente diferentes unas de otras, y esto es lo que cabía esperar, aunque solo fuera porque somos herederos de una larga y compleja tradición ética compuesta a partir de muchas y distintas tendencias religiosas y sociales.

Como empresa que intenta ser descriptiva, tanto como la antropología, el proyecto reduccionista está mal concebido. No obstante, puede tener otros objetivos. Puede aspirar, en algún nivel más profundo, a darnos una teoría del objeto de la ética. Pero tampoco está claro por qué esa meta tiene que alentarnos a reducir nuestras concepciones éticas básicas. Si existe una cosa tal como la verdad referida al objeto de la ética —la verdad, podríamos decir, sobre la ética—, ¿por qué se tiene la expectativa de que sea simple? En particular, ¿por qué habría de ser conceptualmente simple, y usar solo uno o dos conceptos éticos, tales como el de *deber* o el de un *estado de cosas bueno,* en vez de muchos conceptos? Quizá para su descripción nos hagan falta tantos conceptos como nos sean necesarios, no menos.

La razón para tratar de reducir nuestros conceptos éticos tiene que hallarse en un objetivo diferente de la teoría ética, que no es exclusivamente el de describir cómo pensamos sobre lo ético, sino el de decirnos cómo hemos de pensar sobre ello. Más adelante defenderé que la filosofía no debería tratar de construir una teoría ética, aunque esto no significa que no pueda ofrecer ninguna crítica de las creencias e ideas

---

[12] G. E. Moore, *Principia Ethica* (1903) (Cambridge University Press, 1959), secciones 17 y 89 [*Principia Ethica,* trad. de M. Vázquez Guisán (Barcelona, Crítica, 2002)]. Moore rechazó esta concepción en «A Reply to My Critics», en P. A. Schilpp (ed.), *The Philosophy of G. E. Moore* (La Salle, Open Court Publishing Co., 1942). En *Ethics* (1912) ya parece rechazarla, aunque solo dice que se abstiene de afirmarla.

éticas. Sostendré que en ética la empresa reduccionista no tiene justificación y que debería desaparecer. Lo que me interesa aquí, sin embargo, es solo subrayar que el reduccionismo necesita una justificación. Buena parte de la filosofía moral se aferra a esto, y la única razón que proporciona es que así es como ha procedido desde hace mucho tiempo.

Hay un motivo favorable al reduccionismo que no opera simplemente sobre lo ético o sobre lo no ético, sino que tiende a reducir toda consideración a un tipo básico. Esto descansa en un presupuesto sobre la racionalidad, según el cual dos consideraciones no pueden cotejarse racionalmente entre sí a menos que haya una consideración común en términos de la cual puedan compararse. Este presupuesto es tan fuerte como totalmente infundado. Al margen de lo ético, las consideraciones estéticas se pueden cotejar con las económicas (por ejemplo) sin ser una aplicación de estas y sin que sean ambas un ejemplo de un tercer tipo de consideración. Los políticos saben que las consideraciones políticas no están todas hechas del mismo material que otras consideraciones con las que se evalúan; incluso consideraciones políticas diferentes puede estar hechas de materiales diferentes. Si uno compara un empleo, unas vacaciones o un compañero con otro, el juicio no requiere un conjunto específico de instrumentos de evaluación.

No se trata solo de un error intelectual. Si lo fuera, no sobreviviría al hecho de que las personas tienen experiencias que lo contradicen, de que habitualmente llegan a conclusiones que tienen por racionales, o por lo menos razonables, sin el uso de un patrón de comparación. El impulso hacia una *concepción racionalista de la racionalidad* proviene, en cambio, de los rasgos sociales de la vida moderna, que imponen a la deliberación personal y a la misma idea de razón práctica un modelo que deriva de una comprensión particular de la racionalidad pública. Esta comprensión exige que, en principio, toda decisión se base en fundamentos que se puedan explicar de forma argumentada. El requisito de hecho no se satisface, y es probable que contribuya poco a alcanzar el objetivo de que, genuinamente, la autoridad rinda cuentas. Pero es una idea influyente e, invirtiendo el orden de las causas, puede parecer como si fuera el resultado de aplicar al mundo público un ideal independiente de racionalidad. En tanto que ideal, volveremos a él más adelante[13].

---

[13] En determinadas formas, la exigencia de una racionalidad discursiva explícita es tan vieja como Sócrates y no representa ninguna influencia moderna, pero los modelos más poderosos de justificación vigentes actualmente y las exigencias de un único patrón

Regresemos a la pregunta de Sócrates. Se trata de un ejemplo particularmente ambicioso de cuestión personal práctica. En contraste, la pregunta más inmediata y sin complicaciones de ese tipo es «¿Qué voy a hacer?» o «¿Qué haré?». Las diversas consideraciones éticas y no éticas que hemos examinado contribuyen a contestar esta pregunta. Su respuesta, la conclusión de la deliberación, tiene la forma «haré...» o «lo que voy a hacer es...»; y con esto expreso una intención, que me he formado como resultado de mi deliberación. Cuando llega la oportunidad de actuar, puede ser que no la lleve a cabo, pero entonces será porque lo olvide, o me lo impidan, o haya cambiado de idea, o porque (como puedo descubrir) nunca quise realmente hacerlo dado que no fue realmente la conclusión de mi deliberación, o no fue realmente una deliberación. Cuando el momento de actuar es inmediato, hay menos espacio para estas alternativas, por lo que resulta paradójico que llegue a una respuesta de este tipo e inmediatamente no haga lo que dije que iba a hacer inmediatamente.

La pregunta «¿Qué he de hacer?» proporciona más margen de maniobra entre el pensamiento y la acción. En este caso la conclusión apropiada es «he de hacer...» y hay varios modos inteligibles de añadir aquí «...pero no voy a hacerlo». *He de* remite a las razones que tengo para actuar de una manera y no de otra. La función normal de «he de..., pero no voy a hacerlo» es llamar la atención sobre cierta clase especial de razones, tales como las éticas o las prudenciales, que son particularmente buenas como razones para manifestar a los demás —porque sirven para justificar mi conducta, por ejemplo, haciéndola encajar en el plan de acción de alguien—, pero que no son, como sucede en este caso, las razones más poderosas para mí ahora; la razón más poderosa es que deseo ardientemente hacer otra cosa. El deseo de hacer algo es, por supuesto, una razón para hacerlo[14]. (Puede incluso ser una razón que

---

para las razones son sin duda expresiones de la racionalidad burocrática moderna. Esta cuestión está relacionada con la de la historia del «mero» yo, o sujeto sin definición social, al que me referí en la nota 6 anterior; MacIntyre exagera la medida en que esta es una concepción puramente moderna. Sobre la pregunta de qué racionalidad se puede pedir razonablemente a un sistema de decisión, aunque esté expresado en términos formales, véase Amartya Sen, *Collective Choice and Social Welfare* (San Francisco, Holden Day, 1970) y «Rational Fools», reimpreso en *Choice, Welfare and Measurement* (Oxford, Blackwell, 1982); A. Sen y B. Williams (eds.), *Utilitarianism and Beyond* (Nueva York, Cambridge University Press, 1982), Introducción, págs. 16-18.

[14] Esto lo niega Edward J. Bond en *Reason and Value* (Nueva York, Oxford University Press, 1983).

justifique mi conducta ante los demás, aunque hay ciertas tareas justificativas, en particular las conectadas con la justicia, que por sí misma no puede hacer). Por ello, en este tipo de caso lo que creo que es más razonable para mí hacer, considerándolo todo en su conjunto, es lo que deseo ardientemente hacer, y si *he de* se considera que remite a lo que es más razonable para mí hacer, esto es lo que he de hacer. Hay otra cuestión, más profunda: si puedo dejar de hacer de manera intencional y sin compulsión incluso lo que creo que es más razonable para mí hacer, esto se conoce como el problema de la *akrasia,* desde que Aristóteles denominara así al fenómeno[15].

La pregunta de Sócrates significa, entonces, «¿De qué manera es más razonable para uno vivir?». Al decir antes que la fuerza del *he de* en la pregunta era solo *he de,* quería decir que la pregunta no privilegia ningún tipo de razón sobre otro. En particular, no concede ninguna consideración especial a las razones justificadoras respetables. Si, por ejemplo, las razones éticas aparecen de manera significativa en la respuesta, no será porque hayan sido preseleccionadas por la pregunta.

Sin embargo, la pregunta de Sócrates pone un énfasis específico en distanciarse de cualquier ocasión efectiva y particular en la que quepa considerar qué hacer. Es una pregunta general sobre qué hacer, porque pregunta cómo vivir; y es también, en un sentido, una pregunta intemporal, dado que me invita a pensar en mi vida en general y no en un momento específico de su curso. Estos dos hechos la convierten en una pregunta reflexiva. Esto no determina la respuesta, pero sí que la afecta. Al responder a una pregunta práctica en un momento particular, en una situación concreta, estaré particularmente concernido por lo que quiero *entonces.* La pregunta de Sócrates no la hago en ningún momento en concreto, o, mejor dicho, el momento en el que ciertamente me la planteo no guarda una relación particular con la pregunta. Así pues, la pregunta me fuerza a adoptar una perspectiva más general, y a más largo término, sobre la vida. Esto no determina que la responda desde una perspectiva prudencial a largo plazo. La respuesta a la pregunta podría ser: el mejor modo para mí de vivir es hacer en todo momento

---

[15] Para un análisis de la acrasia, véase Donald Davidson, «How Is Weakness of Will Possible?», en Joel Feinberg (ed.), *Moral Concepts* (Nueva York, Oxford University Press, 1969) [en castellano en *Ensayos sobre acciones y sucesos,* trad. de O. Hansberg, J. A. Robles y M. Valdés (Barcelona, Crítica/IIF-UNAM, 1995)]; y David Pears, *Motivated Irrationality* (Nueva York, Oxford University Press, 1984).

lo que prefiero hacer en ese momento. Pero, si ya tengo una tendencia a la prudencia, es probable que la naturaleza de la pregunta de Sócrates la saque a relucir.

Es, además, la pregunta de *cualquiera*. Esto no significa, por supuesto, que cuando la hace una persona en particular sea una pregunta acerca de cualquiera: es una pregunta sobre esa persona en particular. Pero, cuando se me plantea a la manera socrática, invitándome a la reflexión, será parte de la reflexión, porque es parte del conocimiento que la constituye, el hecho de que la pregunta puede planteársela cualquiera. Una vez está constituida de esta manera, es muy natural pasar de la pregunta, planteada por cualquiera, de «¿Cómo he de vivir?» a la de «¿Cómo ha de vivir cualquier persona?». Esta segunda parece preguntar por las razones que todos nosotros compartimos para vivir de una manera y no de otra. Parece preguntar por las condiciones de la *vida buena;* la vida apropiada, quizá, para los seres humanos en cuanto tales.

¿Qué recorrido puede darle la reflexión socrática a la pregunta en esa dirección, y con qué efectos sobre la respuesta? La intemporalidad de la reflexión no determina que en la respuesta prevalezca a la prudencia. Asimismo, el hecho de que la pregunta reflexiva pueda plantearla cualquiera debería permitir que su respuesta fuera egoísta. Pero, si es egoísta, se tratará más de un tipo de egoísmo que de otro: el egoísmo general, especificado antes, según el cual todas las personas han de favorecer sus propios intereses. De modo natural, esto suscita el pensamiento de que, si es así, entonces tiene que ser mejor una vida humana que se vive de esa manera. Pero, en ese caso (es tentador continuar), tiene que ser mejor, en cierto sentido intemporal o interpersonal, que las personas vivan de esa manera. Al haber sido conducidos a este punto de vista impersonal, quizá se nos pueda requerir que miremos hacia atrás desde él, que hagamos el camino en sentido contrario, e incluso que revisemos nuestro punto de partida. Porque, si no es mejor desde un punto de vista intemporal que cada persona viva de un modo egoísta, quizá tengamos razones para decir que cada uno de nosotros no ha de vivir de esa manera y que, después de todo, tenemos que dar una respuesta no egoísta a la pregunta de Sócrates. Si realmente se sigue esta conclusión de la pregunta de Sócrates, entonces el mero hecho de formularla de manera reflexiva nos permitirá avanzar mucho en la dirección del mundo ético. Pero, ¿se sigue esta conclusión?

El pensamiento práctico es fundamentalmente un pensamiento que se formula en primera persona. Tiene que plantear y responder la

pregunta «¿Qué voy a hacer?»[16]. Además, parece que la reflexión socrática nos conduce a generalizar el *yo* y a adoptar, a partir de la mera fuerza de la reflexión, una perspectiva ética. En el capítulo 4 veremos si la reflexión puede llevarnos tan lejos. Pero incluso si no puede hacerlo, la reflexión socrática nos conduce, sin duda, a algún lugar. La reflexión conlleva cierto compromiso, parece, e innegablemente la filosofía está comprometida con la reflexión. Así pues, la mera existencia de este libro tiene que plantear la doble cuestión de en qué medida la reflexión nos compromete y por qué hemos de comprometernos con ella. Sócrates pensaba que su reflexión era ineludible. Lo que quería decir con eso no era que todos se comprometerían con ella, porque sabía bien que no todo el mundo lo haría; ni que todos los que se pusieran a reflexionar sobre su vida se verían forzados a continuar por compulsión interna, incluso contra su voluntad. Lo que pensaba era, más bien, que la reflexión es parte de la bondad de la vida buena: *una vida sin examen,* como él decía, *no es digna de ser vivida.*

Esto exige una respuesta muy especial a su pregunta, la cual, para él, proporciona la justificación última para plantearla en primer lugar. Mi libro se compromete a plantear la cuestión; pero, ¿se compromete a responderla de ese modo? ¿Cualquier tipo de investigación filosófica de lo ético y de la vida buena tiene que exigir que el valor de la filosofía y de una actitud intelectual reflexiva formen parte de la respuesta?

---

[16] Esto no supone olvidarse de «¿qué *haremos?*», que también es una pregunta en primera persona. La cuestión básica es la de a quién toma el hablante como la primera persona del plural; un hablante que, cabe recordar, es un *yo*.

Capítulo 2

El punto arquimédico

Autores que en otros asuntos se muestran tranquilos se impacientean cuando se preguntan por cómo justificar la moral. A menos que la filosofía pueda proporcionar una justificación para la vida ética, o (más concretamente) la moral, estaremos a merced del relativismo, el amoralismo y el desorden. Según su manera usual de expresarlo: cuando un amoralista pone en duda las consideraciones éticas, y señala que no hay razones para cumplir con las exigencias de la moral, *¿qué podemos decirle?*

Pero ¿qué podemos decir al amoralista en el caso de que exista una justificación para la moral? Bien, podemos ponerle la justificación delante de las narices. Pero ¿por qué habríamos de esperar que con esto ceje en su empeño? ¿Por qué habría de escuchar? El amoralista, o incluso su colega más teórico, el relativista, se presenta en los escritos de esos autores como una figura inquietante, como una amenaza. ¿Por qué habría de importarle a una persona así que haya una justificación filosófica de la vida ética?

Una vez por lo menos, en la historia de la filosofía, se ha representado de manera concreta al amoralista como una figura inquietante, en el personaje Calicles que aparece en el *Gorgias* de Platón. Calicles, según las convenciones del diálogo platónico, entabla una conversación racional y se deja humillar por la argumentación de Sócrates (una argumentación tan poco convincente que Platón tuvo que escribir con posterioridad la *República* para mejorarla). Lo que resulta inquietante en

Calicles es, sin embargo, algo que Platón nos ilustra y que es también el tema del diálogo: que desprecia lúcidamente la filosofía, y solo por condescendencia o diversión se queda a escuchar los argumentos de esta.

—No es eso. La cuestión no es si se convencerá, sino si debe convencerse.

—Pero ¿lo es? La impaciencia de los autores apunta a algo más; apunta a que lo que sucederá podría depender del resultado de la argumentación, que la justificación de la vida ética podría ser una *fuerza*. Si vamos a tomarnos esto en serio, entonces la cuestión de quién se supone que escuchará es una cuestión real. ¿Por qué se supone que escucharán? ¿De qué servirá la justificación del profesor cuando echen abajo la puerta, le rompan las gafas y se lo lleven?

En cualquier caso, incluso si hay algo que el resto de nosotros consideraría una justificación de la moral o de la vida ética, ¿es cierto que el amoralista, llamémosle Calicles, *debe* convencerse? ¿Significa solo que sería bueno que se convenciese? Sin duda, lo sería para nosotros, pero esta no es la cuestión. ¿Significa que es bueno para él? ¿Está siendo imprudente, por ejemplo, al actuar contra lo que es mejor para él? O ¿es irracional en un sentido más abstracto, contradiciéndose a sí mismo o yendo contra las reglas de la lógica? Y, si lo es, ¿por qué tiene que estar preocupado por ello? Robert Nozick ha planteado acertadamente la cuestión de qué fuerza tiene la acusación de incoherencia hecha al «hombre inmoral»:

> Supongamos que demostramos que cierta posición X que él mantiene o acepta le compromete a comportarse moralmente. Así, ahora tiene que renunciar, por lo menos, a una de las siguientes cosas: (a) comportarse inmoralmente, (b) mantener X, (c) ser coherente con esta posición desde el punto de vista de la moral. El hombre inmoral respondería: «A decir verdad, si tuviera que elegir, renunciaría a la coherencia»[1].

No es evidente en qué habría de consistir una justificación de la vida ética, o por qué habríamos de necesitarla. Esta supuesta justifica-

[1] Robert Nozick, *Philosophical Explorations* (Cambridge, Mass., Harvard University Press, 1981), pág. 408. [Hay trad. cast. de J. Marqués, *Explicaciones filosóficas* (Madrid, Innisfree, 2014)].

ción plantea tres preguntas: ¿A quién va dirigida? ¿Desde dónde? ¿Contra qué? Comenzaremos por la última cuestión, dado que tenemos que preguntarnos qué se propone como alternativa a la vida ética, y es importante que haya alternativas. «Amoralista» es el nombre de alguien. Esto permite delimitar las preguntas en relación con una inquietud filosófica recurrente, el escepticismo. El escepticismo puede afectar a cualquiera de las cosas que las personas dicen saber: que existe un mundo «externo»; que otras personas tienen experiencias (que hay otras personas, se puede decir también); que la investigación científica puede producir conocimiento; que las consideraciones éticas tienen fuerza. El escepticismo filosófico afecta a todas esas cosas, pero de maneras muy diversas y con efectos muy distintos. En el caso del mundo externo, para una persona cuerda, la cuestión real planteada por el escepticismo no es si alguna cosa de las que decimos sobre el mundo es verdadera, o incluso si sabemos que alguna lo es, sino cómo sabemos que alguna de ellas es verdadera, y en qué medida. La vida no ofrece ninguna alternativa a esas creencias: cualquier alternativa tendría que ser una alternativa a la vida. En el caso de las «otras mentes», como se conoce al problema, se puede decir casi lo mismo dentro de los límites de la cordura, pero el problema vira inquietantemente hacia la cuestión de saber *cuánto sabemos*. Sin duda, sabemos que los demás tienen sentimientos, ¿pero qué grado de conocimiento tenemos de esos sentimientos? Esta es, en parte, una cuestión filosófica, que tiene un efecto más práctico que la mera pregunta «¿cómo lo sé?».

El escepticismo ético, a este respecto, está en el extremo opuesto al escepticismo sobre el mundo externo. No es, por otro lado, como el escepticismo acerca de la investigación psíquica o el psicoanálisis, en los que se plantea una duda real que podría llegar a aceptarse, con el resultado de que esas actividades tendrían el mismo destino que la frenología: las rechazaríamos en su conjunto, al hallar que sus pretensiones de conocimiento o incluso de creencia razonada eran infundadas. No es posible que las consideraciones éticas sean rechazadas colectivamente de esta manera. Sin embargo, para el individuo sí que parece haber una alternativa a aceptar las consideraciones éticas. Se halla en una vida que no es una vida ética.

Este tipo de escepticismo ético difiere tanto del escepticismo sobre el mundo externo que no puede ser tratado con los mismos métodos. Es conocido que Moore desconcertó al escéptico acerca de los objetos materiales confrontándolo con uno de estos objetos, la mano del pro-

pio Moore (en todo caso, habría sido una confrontación si un escéptico de este tipo hubiera estado allí)[2]. Se ha debatido mucho sobre los efectos del gesto de Moore —sobre, por ejemplo, si era una petición de principio—, pero indudablemente tiene ciertos efectos, al recordarnos que tomarnos a este escéptico en serio podría suponer tomar al pie de la letra lo que dice, lo que no dejaría de causar problemas. No hay ninguna analogía aquí con lo ético. Podría ser que si hay verdades éticas, algunas de ellas puedan ser reconocidas como ciertas: si se puede elegir, uno no habría de intervenir quirúrgicamente a un niño sin anestesia, pongamos por caso[3]; pero el efecto perturbador que tiene este ejemplo para el escéptico ético es distinto del que tiene el de la mano de Moore para el otro tipo de escéptico. Para empezar, una proposición aislada que se sabe que es verdadera acerca de un objeto material acabará con este primer tipo de escepticismo: la mano de Moore es un ejemplo de un objeto material, y, como *uno* refuta a *ninguno,* así *ciertamente uno* refuta a *posiblemente ninguno*. Pero el ejemplo del niño, o cualquier otro caso aislado de un tipo tan llamativo, contará como ejemplo de lo ético solo para quien reconoce lo ético. El amoralista, el mismo Calicles si se siente inclinado a hacerlo, podría ayudar o socorrer a un niño. Un sentimiento limitado de benevolencia o de altruismo podría motivar a casi cualquiera a pensar que habría de actuar de cierta manera en una ocasión dada, pero este hecho no lo pone ante lo ético, como la mano de Moore pone al escéptico ante algo material. Lo ético supone mucho más, todo un entramado de consideraciones, y el escéptico ético podría tener una vida que le permitiera ignorar por completo tales consideraciones.

El escepticismo del escéptico tradicional se refería fundamentalmente al *conocimiento,* pero el escéptico ético no es necesariamente alguien que duda de si hay conocimiento ético. Tal como yo lo entiendo, ser escéptico acerca de la ética es ser escéptico acerca de la fuerza de las consideraciones éticas; una persona puede concederles fuerza y no ser

---

[2] G. E. Moore, «Proof of an External World», reimpreso en *Philosophical Papers* (Atlantic Highlands, N. J., Humanities Press, 1977) [«Prueba del mundo externo», en *Defensa del sentido común y otros ensayos,* trad. de C. Solís (Madrid, Taurus, 1972; Barcelona, Orbis, 1983)]. Sobre la fuerza de la respuesta mooreana al escepticismo, véase Thompson Clarke, «The Legacy of Skepticism», *Journal of Philosophy,* 69 (1972).

[3] Renford Bambrough, *Moral Scepticism and Moral Knowledge* (Atlantic Highlands, N. J., Humanities Press, 1979), pág. 15.

escéptica, pero aun así no creer que constituyan conocimiento porque no piensa que su interés resida en el hecho de si son conocimiento. (Para la cuestión de si hay conocimiento ético, véase el capítulo 8). Pero, incluso si el escepticismo ético se entiende de este modo, no deberíamos asumir que el escéptico tiene que ser alguien que lleva una vida que va contra las consideraciones éticas. Quizá haya que decir, más bien, que da cabida a esa vida. Un escéptico, después de todo, es meramente escéptico. En la medida de lo posible ni afirma ni niega, y el escéptico total, el pirrónico de la antigüedad, se suponía que ni afirmaba ni negaba nada. No podría conseguirlo[4], y es dudoso que el escéptico ético pudiera conseguir tal cosa —utilizar el vocabulario ético mientras suspende el juicio con respecto a toda cuestión ética—. Las dificultades residen en la noción misma de un comportamiento de ese tipo. Es difícil, por ejemplo, usar el vocabulario de la promesa y al mismo tiempo sostener la posición de que no hay nada decisivo que decir, a favor o en contra, con respecto a si deben mantenerse las promesas. Además, el escéptico tiene que actuar, y si se incorpora al mundo del discurso ético, entonces lo que hace tiene que tomarse como expresión de pensamientos que obtiene en ese mundo. Si habla de acciones que están bien éticamente o no, y realiza una acción de buen grado, entonces tenemos que pensar que cree que esta está bien. Por ello, esta no es una opción para el escéptico ético. Pero hay otra opción, que consiste en dejar de usar el discurso ético por completo, salvo quizá para engañar. Aunque no es una cosa fácil de hacer, el escéptico podría ser capaz de demostrar que no se ve afectado en absoluto por las consideraciones éticas. Esto pone de relieve la importancia de la existencia de alternativas. El escéptico no se queda sin nada que hacer.

Las motivaciones que puede conservar el amoralista son un elemento *sobre* el cual las pretensiones éticas podrían elaborar una justificación. Aun así, es un error (como veremos en el siguiente capítulo) pensar que hay cierta presunción objetiva a favor de la vida no ética, que el escepticismo ético es el estado natural y que la persona que estamos imaginando es la que todos nosotros quisiéramos ser si la vida ética no tuviese justificación y hubiéramos descubierto que no la tiene. El filó-

---

[4] Véase Myles Burnyeat, «Can the Sceptic Live his Scepticism», en Malcolm Schofield, Myles Burnyeat y Jonathan Barnes (eds.), *Doubt and Dogmatism: Studies in Hellenistic Epistemology* (Nueva York, Oxford University Press, 1980); reimpreso en Myles Burnyeat (ed.), *The Skeptical Tradition* (Berkeley, University of California Press, 1983).

sofo moral que busca justificaciones a veces hace ver que es así, sobrestimando a este respecto la necesidad de una justificación tanto como había sobrestimado su efecto, al menos su efecto sobre el escéptico practicante.

Esto nos hace regresar a la pregunta de «¿A quién?». Cuando el filósofo plantea la cuestión de qué tendríamos que decir al escéptico o al amoralista, haría mejor en preguntar qué tendríamos que decir sobre sus afirmaciones. La justificación que él busca está diseñada para las personas que están en gran medida dentro del mundo ético, y el objetivo de su discurso no es bregar con alguien que probablemente no lo escuchará, sino tranquilizar, reforzar y dar claves interpretativas a quienes quieran formar parte de ese mundo. Esto sitúa en una perspectiva diferente la idea que vimos expuesta de forma bastante optimista para el caso del amoralista, según la cual la justificación del mundo ético sería una fuerza. Platón, quien vio más profundamente que ningún otro filósofo las cuestiones planteadas por la posibilidad de una vida fuera de la esfera de lo ético, no dio por sentado que una justificación de la vida ética pudiese ser una fuerza. Pensó que el poder de la ética era el poder de la razón, y que esta tenía que ser *transformada* en una fuerza. Según su parecer, se trataba de un problema político, y así es. Pero creyó que la justificación era intelectual y muy difícil de establecer y, además, que todos teníamos cierta inclinación natural a escapar al orden ético y destruirlo. Esta inclinación era una presencia constante en la mayoría de personas, que carecían de la capacidad para dominar la justificación y, con ello, dominarse a sí mismos[5]. Para Platón, el problema político de transformar lo ético en una fuerza era el problema de hacer que la sociedad asumiera la justificación racional, y el problema solo podía tener una solución autoritaria. Si, por el contrario, la justificación se dirige a una comunidad que ya es ética, entonces la política del discurso ético, incluida la filosofía moral, es significativamente diferente. El objetivo no es controlar a los enemigos de la comunidad o a quienes se zafan, sino dar razones a las personas que ya están dispuestas a escuchar para contribuir a la creación de una comunidad que se mantenga unida por esa misma disposición.

---

[5] Cfr. «The Analogy of City and Soul in Plato's Republic», en E. N. Lee, A. P. Mourelatos y R. M. Rorty (eds.), *Exegesis and Argument. Studies in Greek Philosophy presented to Gregory Vlastos* (Assen, Van Gorcum, 1973).

Hasta el momento he estado suponiendo casi siempre que si podemos debatir racionalmente con alguien, entonces nosotros y esa persona vivimos una forma de vida ética (aunque no necesariamente de la misma): es improbable que las personas que viven al margen de toda vida ética intercambien argumentos con nosotros, y tenemos pocas razones para confiar en ellas si lo hacen. Pero podría ser de otra forma. Dejando de lado los motivos inconexos e incluso artificiales que tenía Calicles para conversar con Sócrates, lo importante es que las personas pueden verse movidas por una necesidad común —en el caso límite, por un temor común al desastre— a negociar acuerdos de cooperación restringida o, por lo menos, de no agresión. Hay razones inherentes por las que tales acuerdos, sin cierta sanción externa, están destinados a ser inestables[6]. En cualquier caso, no generan ninguna comprensión ética compartida. Esto es suficiente para mostrar que las personas pueden debatir racionalmente sin compartir un sistema ético. Tal vez, para un propósito limitado, podrían debatir racionalmente sin que ninguna de ellas tuviera un sistema ético. Aun así, por lo general esto no es posible, pues para que una discusión racional se dé efectivamente entre dos partes, se precisa de algo que la sostenga. Por supuesto, puede ser una relación particular que generalmente no vaya más allá de lo ético, pero, si no es así, si no nos encontramos ante la condescendencia de Calicles o ante la necesidad que une a la gente en un caso de emergencia, entonces esta relación tiene que incorporar un mínimo de consciencia ética.

Esto pone de relieve de nuevo el tópico de que es imposible que todos los miembros de una comunidad vivan fuera de la vida ética, aunque sí puede hacerlo una persona sola. Lo que nos lleva a formular en primera persona la pregunta de si es posible justificar las consideraciones éticas a partir de cero. Un agente que se plantee la pregunta de Sócrates puede preguntarse si podría llegar a tener razones para llevar una vida ética, presuponiéndole una estructura mínima de acción, deseo o creencia. Este agente no tiene por qué ser alguien que posee solo las características mínimas: no tiene por qué estar fuera del mundo

---

[6] Puede encontrarse una formulación precisa de estas inestabilidades en el problema del dilema del prisionero. Véase, por ejemplo, Roberto D. Luce y Howard Raiffa, *Games and Decisions* (Nueva York, John Wiley, 1957). La teoría política de Hobbes se basa en una manera determinada de afrontar este problema. Sobre la relevancia de estas cuestiones para la ética, véase Edna Ullmann-Margalit, *The Emergence of Norms* (Nueva York, Oxford University Press, 1977).

ético, tratando de ver si hay una manera de entrar en él. Aún mejor, puede formar parte de él y preguntarse qué razones tiene para estar ahí. (De nuevo, los resultados de su reflexión influirán en su forma de entender su misma reflexión). Aquí no tenemos ningún problema con respecto a la pregunta «¿A quién?». La que importa ahora es la restante de nuestras tres preguntas iniciales, «¿Desde dónde?». ¿Cuáles con las características mínimas que se supone que tiene esta persona? Si está tratando de justificar la vida ética desde la base, ¿dónde está la base?

Por retomar una imagen muy manida, ¿dónde se encuentra el punto arquimédico? La pregunta no está solo manida, sino que también desincentiva profundamente cualquier indagación a la que pueda representar. En el caso de ciertas indagaciones, nos resulta tan familiar el desánimo que vemos difícil poder imaginar en qué consistiría el éxito. Si se nos asignara la tarea de encontrar una posición desde la que validarlas, con independencia de todos nuestros conocimientos y creencias, no podríamos comprender la idea lo suficiente como para reconocerla como un cometido. En el caso de la ética, sin embargo, tenemos una idea más clara de en qué consistiría este cometido. Nos haría falta encontrar un punto de apoyo en la idea de acción racional. Como hemos visto, esa idea, de por sí, no supone un compromiso inmediato con la ética. A ello se debe que la pregunta de Sócrates no sea desde el principio una pregunta ética, y es también por lo que el amoralista o el escéptico parecen sostener la posibilidad de una vida racional al margen de las consideraciones éticas. Con todo, quizá esto solo sea una primera impresión previa a una reflexión más pausada. La pregunta socrática no hacía uso de ningún término específicamente ético, y esto sigue siendo así. No obstante, podría resultar que si reflexionamos correctamente sobre ello descubramos que estamos comprometidos con una vida ética tan solo por ser agentes racionales. Algunos filósofos creen que esto es verdad. Si tienen razón estos filósofos, entonces existe lo que he llamado un punto arquimédico: algo con lo que está comprometido incluso el amoralista o el escéptico, pero que, bien ponderado, nos mostrará que son irracionales, o poco razonables, o en todo caso que están equivocados.

Hay dos clases básicas de aventura filosófica que responden a este patrón. Una de ellas parte de la concepción más mínima y más abstracta posible de la agencia racional. La examinaremos en el capítulo 4. La otra, de la que nos ocuparemos de inmediato, defiende una concepción más rica y más determinada de lo que es la agencia racional, a la que

considera expresión de un modo específico de vida humana. Ambos planteamientos están enraizados en filosofías del pasado; la concepción más rica y más determinada, proviene de Aristóteles; la más abstracta, de Kant. Para ambas filosofías, la pregunta de Sócrates no puede entenderse sencillamente como es presentada, como una pregunta, por así decir, que espera una respuesta. Las dos redefinen la búsqueda de un punto arquimédico y lo hacen de maneras distintas. Pero tienen algo importante en común que las conecta con el cuestionamiento inicial de Sócrates, en contraste con otras líneas menos fructíferas de la historia de la filosofía moral. Ambas proporcionan una argumentación propia de la razón práctica. Ninguna aspira, en primera instancia, a demostrar la verdad de ciertas proposiciones éticas, que tendríamos que aceptar en virtud de nuestro interés por creer en la verdad. Por el contrario, nos recomiendan determinadas acciones por nuestro interés en actuar racionalmente o en llevar una vida humana satisfactoria. Tanto Aristóteles como Kant entienden que la justificación de las proposiciones éticas procederá solo de esto: de que son proposiciones aceptadas por quien actúa racionalmente o lleva una vida satisfactoria.

Capítulo 3

Los fundamentos: el bienestar

Sócrates plantea su pregunta, en la *República* de Platón, en el transcurso de una discusión con Trasímaco, un personaje de ficción creado, al parecer, para encarnar algunas de las tesis más duras de sus contemporáneos los sofistas. Trasímaco concede que a menudo existen razones para preocuparse por los intereses de los demás tanto como por los propios, pero mantiene que esto se debe solo a que nuestro poder está limitado, por lo general, por el poder superior de los demás. Según Trasímaco, los seres humanos buscan naturalmente el poder y el placer, aunque pueden tener que restringir, de un modo racional, esa búsqueda ante el poder de otras personas. También pueden llegar a pensar, irracionalmente, que es correcto o noble respetar los intereses de los demás, pero en tal caso se equivocan por culpa de las convenciones, las reglas sociales que inculcan esos presupuestos respetables pero infundados. Si piensan así, suele ser porque otra persona tiene un poder superior; su error es un engaño, y las convenciones que les embaucan son un instrumento de coacción.

Trasímaco dice que las convenciones que exhortan a respetar los intereses ajenos —la «justicia»[1], como pueden llamarse— son un ins-

[1] La palabra griega *dikaiosyne,* de la que esta es la traducción, tiene un sentido más amplio en el uso que de ella hace Platón que nuestra expresión equivalente. He dado una

trumento de los fuertes para explotar a los débiles. Esto plantea inmediatamente la pregunta de qué hace fuertes a esas personas. Trasímaco habla como si el poder político o social no fuera en sí mismo convencional, y esta es una concepción que es difícil que sea apta para un patio de escuela. En la *República,* inmediatamente después se presenta otra posición según la cual la justicia es el producto de una convención adoptada por un grupo de personas para protegerse a sí mismas: es un mecanismo contractual que usan los débiles para hacerse fuertes. Esta formulación es, a simple vista, la opuesta de la precedente, y ciertamente es más sofisticada, pero ambas tienen mucho en común. Ambas concepciones entienden la justicia como un instrumento para satisfacer los deseos egoístas que existen naturalmente, con independencia de todo punto de vista ético. Ambas ven la justicia como algo a lo que uno no se plegaría si no le hiciera falta.

Para Platón, esta era una debilidad básica. Pensaba que una caracterización de la vida ética podría responder a la pregunta de Sócrates, y combatir el escepticismo, con demostrar únicamente que era racional para las personas ser justas con independencia de quiénes fuesen y de sus circunstancias. A este respecto, la segunda caracterización, contractual, no era mejor que la brutal versión original. Para un hombre que fuera poderoso e inteligente y que tuviese la suerte de estar bien situado, no sería racional amoldarse a las exigencias convencionales de la justicia. La teoría contractual era particularmente débil porque era desestabilizada por la presencia de un agente superior, más inteligente, capaz y persuasivo que el resto. Era sobre todo a causa de este tipo de agente por lo que Platón pensó que había que luchar contra el escepticismo y demostrar que la justicia y la vida ética son racionales.

Desde este punto de vista, para Platón la teoría contractual fallaba. Y lo hacía, además, a causa de un determinado rasgo estructural: presentaba como éticamente básica una *práctica* deseable o útil, a saber, las convenciones de la justicia. Pero tanto para Platón como para Sócrates, lo éticamente deseable tendría que ser primordialmente algo que se hallase en el agente. Para ellos, si algo exterior al alma (según su lenguaje) es éticamente primario —una regla, por ejemplo, o una institu-

---

explicación más completa de algunas concepciones antiguas sobre estos temas en un artículo mío sobre la filosofía griega que aparece en M. I. Finley (ed.), *The Legacy of Greece* (Nueva York, Oxford University Press, 1981), que se solapa parcialmente con la discusión presente.

ción—, nos exponemos a que pueda haber una persona cuyas necesidades más profundas y el estado de su alma sean tales que le parezca irracional actuar de acuerdo con esa regla o institución y, en tanto que esta eventualidad sea posible, la tarea de responder a la pregunta de Sócrates de un modo favorable a la vida ética no se podría llevar a cabo. La exigencia de mostrar a *cada* persona que la justicia es racional *para esa persona* supone que la respuesta tiene que fundarse, en primera instancia, en dar cuenta de la clase de persona que es racional ser.

Dicen a veces los críticos modernos de la ética platónica —y se objeta lo mismo a la posición de Aristóteles— que es egoísta en un sentido que entra en conflicto con el carácter fundamental de la moral[2]. Los griegos, se afirma, no habían llegado a una comprensión madura de la consciencia moral. No habían llegado, ciertamente, a las preocupaciones que son distintivas del sistema de la *moral,* con su énfasis en una noción muy especial de obligación. (En esto, como veremos después, fueron muy afortunados). Pero ni Platón ni Aristóteles concibieron la vida ética como un medio para incrementar las satisfacciones egoístas. Su planteamiento es formalmente egoísta, pues asume que hay que mostrar a cada persona[3] que tiene buenas razones para vivir éticamente, y estas buenas razones apelan a algo relativo a esa persona, a cómo y qué será si es una persona con ese tipo de carácter. Pero su perspectiva no es egoísta en el sentido de que traten de mostrar que la vida ética está al servicio de cierto conjunto de satisfacciones individuales que están claramente definidas antes de que entren en juego las consideraciones éticas. Su objetivo no es, dada una caracterización del yo y de qué cosas lo satisfacen, mostrar cómo la vida ética (con suerte) contribuye a su realización, sino proporcionar una descripción del yo que concuerde con esa forma de vida.

Este objetivo es ya mucho más sofisticado que el de los planteamientos religiosos más primitivos, que presentan las consideraciones éticas como un conjunto de leyes o mandamientos sancionados por la promesa del castigo o la recompensa divinos. Este nivel más primitivo

---

[2] El presupuesto de que las ideas éticas griegas son inferiores a las modernas, y en particular a las kantianas, es un rasgo desafortunado del conocido libro de A. W. Adkins, *Merit and Responsibility* (Chicago, University of Chicago Press, 1975).

[3] La tesis de Aristóteles es en realidad más débil; véase más adelante en este capítulo.

de la moral religiosa es más egoísta. Sin embargo, no habría que descartar tampoco este tipo de explicación bajo el pretexto de que las motivaciones egoístas que invoca no pueden ser tomadas en consideración, ni porque supuestamente no puede derivarse un *debe* del poder de Dios, sino solo de su bondad[4]. No hay nada incorrecto en la forma general de esta caracterización: explica por qué una persona tendría buenas razones para vivir el tipo de vida que respeta los intereses de los demás. Se trata, más bien, de que sabemos que esto no podría ser verdad —no *podría* ser verdad, porque, si algo entendemos sobre el mundo, es que no funciona así—. De hecho, muchas personas, incluidos muchos cristianos, dirían hoy en día que sabemos que no funciona así.

Es natural decir que esta teoría religiosa es simplista; no porque sea religiosa, sino porque es un aspecto simplista de la religión. Una ética religiosa menos simplista no añadirá el elemento religioso como una mera sanción externa, sino que ofrecerá una caracterización de la naturaleza humana que tendrá en cuenta tanto los objetivos éticos como la relación con Dios. Sin embargo, la crítica de que una teoría religiosa simplista plantea una cuestión importante: si la religión es, en último término, cuestión de cómo es el mundo, ¿por qué el mundo *no* ha de ser tan simplista? ¿Por qué hay que juzgar a la religión en términos de determinadas comprensiones éticas que el lenguaje secular clasifica como más o menos sofisticadas o maduras? La respuesta tiene que ser, presumiblemente, que la idea simplista original de Dios como el guardián omnipotente de la ley se alcanzó por medio de nuestra comprensión ética (simplista). Pero entonces, para que la comprensión ética se desarrolle y para que la religión comprenda su propio desarrollo en relación con eso, parece inevitable que tenga que llegar a comprenderse a sí misma como una construcción humana; pero si lo hace, acabará por desplomarse.

Es verdad que el desarrollo de la consciencia ética supone el desplome de la religión, pero no porque a una ética religiosa, aunque sea simplista, le resulte lógicamente imposible ser ética. Es más bien por la razón dialéctica de que, si la autocomprensión de la religión ha de ser preservada por la consciencia ética, esta autocomprensión ha de tomar

---

[4] Para ciertos trabajos recientes, véase Paul Helm (ed.), *Divine Commands and Morality* (Nueva York, Oxford University Press, 1981). Sobre el egoísmo, véase el capítulo «Dios, moralidad y prudencia», de mi *Introducción a la ética*. Sobre la derivación del *debe* a partir del *es*, véase el capítulo 7.

una dirección que destruirá la religión. El meollo de la cuestión no está en problemas puramente lógicos. De hecho, las cuestiones lógicas o estructurales de la ética religiosa, como muchas cuestiones sobre Dios, son interesantes solo si uno cree en Dios. Si Dios existe, entonces los argumentos relativos a él son argumentos sobre el cosmos y de importancia cósmica, pero, si no existe, no remiten a nada. En este caso, las cuestiones importantes tienen que versar sobre los seres humanos, y por qué, por ejemplo, han creído que Dios existía. Las cuestiones referentes a la ética religiosa son cuestiones acerca de los impulsos humanos que se expresan en ella, y han de abordarse en estos términos. Para quienes no creen en una ética religiosa, es una maniobra de evasión seguir argumentando sobre su estructura: esto desvía la atención de la cuestión importante de qué nos enseñan estas perspectivas sobre la humanidad. Se puede interpretar la frase de Nietzsche «Dios ha muerto» como una invitación a tratar a Dios como a un muerto: hemos de repartir la herencia que deja e intentar escribirle una biografía honesta.

La intención de Platón, volviendo a nuestro tema, era dar una imagen del yo de tal tipo que, si las personas comprendieran adecuadamente lo que eran, se apercibirían de que una vida de justicia no es un bien externo al yo, sino, más bien, un objetivo racional. Para él, como para Aristóteles, si fuera racional perseguir cierto tipo de vida o ser cierta clase de persona, entonces aquellas cosas habían de contribuir a crear un estado de satisfacción llamado *eudaimonía*. Este término se suele traducir como «felicidad», pero, en manos de esos filósofos, su significado difiere del de las concepciones modernas de la felicidad. Hoy en día tiene sentido decir que estás feliz un día, infeliz otro, mientras que la *eudaimonía* concernía a la configuración de toda una vida. Usaré la expresión *bienestar* para este estado.

Sócrates lo planteó en términos de conocimiento y del poder de la razón discursiva, y pudo hacerlo gracias a su concepción drásticamente dualista del alma y el cuerpo. El bienestar era el estado deseable del alma —y esto significa que una persona es un alma, dado que un alma indestructible e inmaterial es lo que uno esencialmente es—[5]. Esto

[5] «¿Y de qué modo te enterraremos?», dijo Critón. «Como queráis», dijo Sócrates «siempre que me atrapéis y no me escape de vosotros». Platón, *Fedón* 115c-d [en traducción de C. García Gual (Madrid, Gredos, 1986)].

era lo que subyacía a la concepción que tenía Sócrates de nuestros intereses más profundos, y le hizo más fácil creer que, como afirmó en una frase famosa, *no existe mal alguno para el hombre bueno** dado que lo único que podría afectarle sería algo que vulnerara el buen estado de su alma, y esta era inviolable. El problema, con este punto de vista, es que la descripción de las motivaciones éticas presenta los intereses personales bajo una luz muy espiritual, cuando el objeto de la ética requiere concebirlos de un modo menos espiritual. Si el daño corporal no es realmente un daño, ¿por qué la virtud nos exige tan imperativamente que no dañemos los cuerpos de otras personas?

Hay otra consecuencia especial que se desprende de la presentación que hace Sócrates, relativa a las esperanzas que este tenía en el poder regenerativo de la filosofía. Él, y en alguna medida Platón, creían que la disciplina de la filosofía, gracias a su poder para desarrollar las virtudes, no podía más que conducirnos al bienestar. La filosofía racional había de proporcionar la perspicacia que permite acceder al bienestar. Esto significaba que la filosofía o bien enseña los medios para satisfacer las necesidades innatas, o bien nos permite redefinir racionalmente nuestras necesidades. El proyecto de Aristóteles es menos ambicioso, y esta es una de las razones, junto con su mayor elaboración psicológica y social, por las que la *Ética*[6] de Aristóteles sirve aún de paradigma para cualquier posición que intente basar la ética en consideraciones relativas al bienestar y a una vida digna de ser vivida. Para él, un ser humano no es un alma inmaterial, sino que es un ser esencialmente encarnado y que lleva una vida social. Aristóteles traza una distinción básica entre los poderes de la razón, de modo que la facultad intelectual central para la vida ética, la razón práctica, tiene funciones y objetos muy diferentes de los de la razón teórica, que se despliega en la filosofía y en las ciencias. Aristóteles pensaba que el cultivo de la filosofía y de las ciencias era la

---

* *Apología* 41c. *[N. del T.]*.

[6] Hay dos obras atribuidas a Aristóteles que llevan la palabra «ética» en su título —la *Ética nicomáquea* y la *Ética eudemia*—, con algunos pasajes comunes a ambas. La concepción establecida es que la dedicada a Nicomáco tiene más visos de autenticidad, pero véase Anthony Kenny, *The Aristotelian Ethics* (Nueva York, Oxford University Press, 1978). Un comentario inestimable es el de Sarah Broadie, *Ethics with Aristotle* (Oxford University Press, 1990). Amelie Rorty (ed.), *Essays on Aristotle's Ethics* (Berkeley, University of California Press, 1981), es una compilación útil.

forma superior de actividad humana, pero presuponía que el ejercicio de la razón práctica en una vida personal y cívica era necesario para conseguirla, no solo en el sentido (platónico) de que tales actividades son necesarias para la sociedad, sino también en el de que todo individuo necesita ese género de vida. Sin embargo, el énfasis oscila en Aristóteles entre la vida cívica como una necesidad que el sabio no puede eludir y (lo que es ciertamente la consecuencia más coherente y más convincente de su antropología filosófica) la vida cívica como una necesidad de todo hombre que desee expresar plenamente sus capacidades.

Para la vida de la razón práctica, son fundamentales determinadas excelencias de carácter o virtudes, que no son otra cosa que disposiciones interiorizadas a actuar, desear y sentir. Por una parte, la teoría de las virtudes de Aristóteles con respecto a la valentía, por ejemplo, o al autocontrol, nos resulta muy familiar; pero en otros aspectos pertenece a otro mundo. Lo importante para la filosofía moral es saber si los elementos más específicos de una cultura pueden separarse de la estructura principal. Esto es posible para algunos de ellos, entre los que se encuentran cuestiones muy importantes; adoptar un planteamiento aristotélico no implica compartir la concepción aristotélica de la esclavitud o de la condición de las mujeres. Sin embargo, la duda surge del hecho de que cuanto más alejado parece Aristóteles de las percepciones éticas modernas, suele ser porque el agente admirado manifiesta una inquietante propensión a interesarse en sí mismo. Aristóteles admite que el hombre bueno necesita amigos, y que la amistad forma parte de la vida buena; pero ve necesario argumentar a favor de esto para reconciliar la amistad con el ideal de autosuficiencia[7]. Incluso su concepción de la veracidad comparte el extraño rasgo egocéntrico de que los vicios y las faltas que opone a esta virtud no son, como sería de esperar, la doblez (la falta de fiabilidad con respecto a la verdad), sino la jactancia y la falsa modestia. Consideraremos más adelante si los aspectos más

---

[7] Aristóteles dice que el amigo del hombre bueno es «otro yo» (*Ética a Nicómaco,* 1166a 31), unas palabras que expresan las tensiones genuinas que alberga su pensamiento entre la amistad y la autosuficiencia. Sin embargo, ahora creo que la crítica de la posición de Aristóteles que hice en otro lugar —véase *Moral Luck* (Nueva York, Cambridge University Press, 1981), pág. 15 [*La fortuna moral,* trad. de S. Marín (México, UNAM, 1993)]— poniendo énfasis en esa expresión era exagerada. Estoy en deuda con Martha Nussbaum con respecto a esta cuestión; véase su libro *La fragilidad del bien,* parte 3. Véase también John Cooper, «Aristotle on Friendship», en Rorty, *Essays on Aristotle's Ethics.*

egocéntricos presentes en la ética de Aristóteles derivan de su propia estructura.

He dicho que, para Aristóteles, una virtud era una disposición interiorizada a actuar, desear y sentir. Es una disposición inteligente. Implica que el agente utilice su facultad de juicio, la cualidad propia de la razón práctica, y por ello no se trata solo de un hábito. También conlleva reacciones favorables o desfavorables hacia los demás, sus caracteres y sus acciones. Las opiniones de Aristóteles sobre este tema están vinculadas a una de las partes más célebres y menos útiles de su sistema, la doctrina del *justo medio,* según la cual cada virtud del carácter se sitúa entre dos faltas o vicios correlativos (como ilustra el ejemplo del respeto a la verdad), que consisten respectivamente en el exceso y la deficiencia de aquello de lo que la virtud representa la cantidad exacta. La teoría oscila entre un modelo analítico inútil (que Aristóteles mismo no sigue de modo consistente) y una doctrina sustantivamente deprimente a favor de la moderación. Es mejor olvidar la doctrina del justo medio, si bien sí implica acertadamente que las personas virtuosas, dado que se supone que saben lo que hacen, verán los fallos o los vicios de los demás como tales y considerarán a quienes los tienen, o por lo menos a sus acciones, como malos, desagradables, poco serviciales o abyectos en distintos grados.

Algunos de nosotros nos resistimos a la idea de que tener una virtud o disposición de carácter admirable haya de implicar también la disposición a juzgar a los demás. Esta resistencia tiene orígenes diversos, algunos de las cuales no tienen nada que ver con la problemática aristotélica. Uno de ellos es una concepción de la inocencia, la imagen de una virtud completamente inconsciente y exenta de la comparación con uno mismo que está implícita cuando juzgamos a otras personas. Otro origen es el escepticismo, la sospecha de que nadie conoce nunca lo suficiente a nadie (incluido, en sus versiones más insidiosas, uno mismo) para poder juzgar. Otro punto de origen se encuentra en el hecho de que aceptamos —y la consideramos una perogrullada— una idea que Aristóteles rechazaba, según la cual se puede tener una virtud sin tener otras. Para Aristóteles, como para Sócrates, la razón práctica exige la armonía entre las disposiciones a actuar y a sentir; para que una disposición pueda contar como virtud, tiene que formar parte de una estructura racional que incluya todas las virtudes. Esto difiere bastante de nuestro presupuesto de que estos tipos de disposición son tan semejantes a otras características psicológicas que explican por qué una persona

puede ser mejor que otra en un determinado ámbito, por ejemplo. Este presupuesto contribuye, además, a inhibir nuestras reacciones para con otras personas.

Al margen de esas consideraciones, sigue habiendo una conexión entre las disposiciones éticas y las reacciones hacia los demás. La naturaleza exacta o la profundidad de esas reacciones, y la garantía que ofrecen, variará en función de los individuos y de las particularidades culturales. Pero, como afirmaba Aristóteles, una disposición ética no es solo un patrón personal de conducta que puede estar contingentemente unido a una tendencia a deplorar o lamentar su ausencia en los demás. Es un tipo de disposición que estructura nuestras reacciones hacia los otros. Debido a que no creemos en la unidad de las virtudes, podemos aceptar la idea de que carecer de determinadas disposiciones éticas es simplemente una peculiaridad de algunas personas. Quizá alguien incluso considere que todas las disposiciones éticas son así, y que ninguna de ellas es tan básica como para impedir que se la califique, por así decir, de anecdótica. Pero alguien que conciba toda disposición ética (o, más bien, aquello sobre lo que trata, dado que él mismo podría no utilizar esos conceptos) como una anécdota carecerá, con seguridad, de toda disposición ética básica. (Huelga decir que puede ser difícil descubrir si tal persona efectivamente concibe así las disposiciones éticas).

Me he referido principalmente a las «reacciones» hacia los demás implicadas en la posesión de disposiciones éticas[8]. Es un término práctico por su amplia y poco concluyente aplicación; hay mucho que decir sobre la gama de las actitudes, tanto positivas como negativas, que puede abarcar esta categoría. Y resulta sorprendente lo poco que ha abordado esta cuestión la filosofía moral, al menos en la tradición anglófona. La razón más importante de esta insuficiencia es el dominio de la moral, que tiende a clasificar todas las reacciones que la incumben —es decir, «morales»— con nombres tales como *juicio, evaluación* y *aprobación* o *desaprobación*. Esto induce a error de diferentes maneras: en primer lugar, todas estas nociones sugieren una posición de superioridad, siquiera temporal, la posición de juez, y es así aunque se den en el

---

[8] Véase P. F. Strawson, «Freedom and Resentment», reimpreso en *Freedom and Resentment and Other Essays* (Nueva York, Methuen, 1976) [«Libertad y resentimiento», incluido en *Libertad y resentimiento y otros ensayos*, trad. de J. J. Acero (Barcelona, Paidós-ICE, 1995)]. Examino las peculiaridades de la moral, y su concepción de la censura, en el capítulo 10.

contexto de una teoría que no aliente esa superioridad. Asimismo, sugieren cierto juicio maniqueo, como podría ser el de culpabilidad o inocencia. Además, se circunscriben, presuntamente, a lo voluntario: nadie puede ser objeto de juicio moral por lo que no es su falta. Al tratar de aferrarse, de este modo, a una justicia última, la moral no se limita a proporcionar una tipología de las reacciones. No está interesada simplemente en qué reacciones cabe denominar morales. La justicia, que es el objetivo de la moral, va más allá de la cuestión de cómo hemos de llamar a nuestras reacciones; pretende solventar el problema de qué reacciones podemos tener en justicia, de modo que primero exige poseer una voz, después la supremacía y, por último, la ubicuidad. Las reacciones «no morales» como el desagrado, el resentimiento o el desprecio, o manifestaciones menores de la vida ética, tales como que alguien sea repulsivo, son encauzadas por una conciencia moral bien educada hacia una retirada rencorosa desde la que urdir la impostura y la venganza.

Esas diversas características del sistema de juicio moral se apoyan mutuamente y se modelan colectivamente a la luz de las prerrogativas de un Dios pelagiano. La severidad de los criterios del juicio responde a la supuesta inmensidad de lo otorgado, la irrevocabilidad de la única justicia última que existe. Por la misma razón, estos criterios invitan, en conjunto, al escepticismo al que he aludido. Se plantea el problema de cómo el carácter y las disposiciones de las personas podrían llegar a ser el objeto de tales juicios. Es muy improbable que las personas sean completamente responsables de su carácter y de sus disposiciones, y es aún menos probable que podamos saber en qué medida lo son (incluso si presuponemos que comprendemos lo que sabríamos si lo supiéramos). Pero ¿nos exige la moral que juzguemos las acciones de las personas al margen de su carácter?

Estas no son preocupaciones aristotélicas. De hecho, Aristóteles pensaba que los seres humanos eran libres en cierto sentido absoluto, y que son los generadores de sus acciones «como lo son de sus hijos»[9]. También pensaba que había reacciones de elogio y censura dirigidas a las acciones y, por medio de las acciones, a las personas, que requerían la voluntariedad de las acciones. Pero no habría podido entender

[9] *Ética a Nicómaco*, 1113b 18. Sobre Aristóteles y el libre albedrío, véase Richard Sorabji, *Necessity, Cause and Blame: Perspectives on Aristotle's Theory* (Ithaca, Cornell University Press, 1980), esp. parte 5.

la insinuación de que esto fija el límite de las reacciones apropiadas a las acciones de los demás habida cuenta de sus disposiciones éticas. Al aceptar, como habríamos de hacer, la idea de que las disposiciones son también disposiciones a reaccionar de una determinada manera, tenemos que recordar que la gama de reacciones es mucho más amplia que la sugerida por las concepciones de la moral.

Aristóteles no habría de haber creído que, por los menos en los aspectos más básicos, las personas eran responsables de su carácter[10]. Su explicación del desarrollo moral en términos de habituación e interiorización deja poco espacio para que la razón práctica pueda alterar radicalmente los objetivos que una persona adulta ya ha adquirido. Aun concediendo esta conclusión, surge un problema relativo a la manera en que Aristóteles presenta su investigación. El problema consiste en determinar cómo concebía la ética filosófica. Él la presenta como una investigación práctica dirigida efectivamente a dar respuesta a la pregunta de Sócrates. Parece como si pensara que podemos revisar la totalidad de nuestra vida y sopesar si responde a los objetivos más valiosos cuando, según su propia teoría, esta no puede ser una representación razonable. Aristóteles concuerda con Platón en que, si la virtud forma parte del bien humano, entonces no puede ser externa al estado de bienestar deseable en último término: ese estado tiene que estar constituido en parte por la vida virtuosa. Sin embargo, esta no es una consideración que pueda usarse para producir ningún efecto radical en el razonamiento práctico, como parece afirmar Aristóteles. Se llega a ser virtuoso, o no se consigue serlo, solo por medio de la adquisición de ciertos hábitos. No deberíamos estudiar filosofía moral hasta que alcancemos la mediana edad, cree Aristóteles, por una razón sintomática de las dificultades que estamos examinando: solo entonces una persona es apta para la deliberación práctica. Pero para entonces ya habrá pasado mucho tiempo desde que se convirtió, con respecto a esta deliberación, en preferentemente bueno o en incorregiblemente malo. (Lo que está aquí en cuestión son solamente los poderes de la razón práctica, pues esto es

---

[10] Aunque Aristóteles dice que lo son, y destaca los casos en los que las personas arruinan su carácter por sus malas prácticas (*Ética a Nicómaco,* 1114a 3-8, 1114b 25-1115a 2), en un pasaje posterior admite que las acciones y los estados del carácter no son «voluntarios del mismo modo». Véase Myles Burnyeat, «Aristotle on Learning to Be Good», en Rorty (ed.), *Essays in Aristotle's Ethics,* que analiza la naturaleza de la indagación de Aristóteles.

del todo coherente con lo que dice Aristóteles de que la vida de una persona puede ser radicalmente modificada por otros medios, tales como la conversión).

Algunos de los razonamientos de Aristóteles podrían tener un efecto deliberativo real. Resulta excelente su análisis de las personas que hacen del honor el objetivo de su vida política, las cuales suelen acabar derrotadas por sí mismas al ponerse en manos de aquellos respecto a quienes aspiran a ser superiores. Para una persona experimentada, este descubrimiento de la *paradoja de Coriolano,* como podríamos denominarla, representaría con toda seguridad una revelación o un diagnóstico. Pero, por regla general, Aristóteles no puede creer, de un modo razonable, que sus reflexiones sobre la vida virtuosa y su contribución al bienestar puedan ayudar a la formación de una deliberación general que podría realizar una persona dada. A la luz de este hecho ha de revisarse la definición de la filosofía ética y de sus aspiraciones. Esta ya no dirige sus consideraciones a cada persona en particular como ayuda para responder la pregunta de Sócrates. Volvemos así a un punto que habíamos adelantado: la respuesta que se dé al escéptico, ante todo, nos beneficiará al resto de nosotros. Aristóteles no está, en realidad, interesado en el escepticismo acerca de la vida ética, y esta es una de las muchas diferencias primordiales entre su mundo y el de Sócrates y Platón. Está preocupado por los hombres que tienen valores equivocados o un mal carácter. Pero el desafío es el mismo: la respuesta a la pregunta de Sócrates no puede ser utilizada por quienes (desde la perspectiva de los demás) más la necesitan.

Aun así, esto no nos aboca al extremo opuesto, esto es, a que la respuesta solo sirve para levantar el ánimo de quienes ya están comprometidos con el sistema, proporcionarles una comprensión más perspicaz y ayudarles a educar a sus hijos. La respuesta sirve para eso, pero no solo para eso. Según la caracterización que hace Aristóteles, la vida virtuosa conduciría al bienestar del hombre que ha tenido una mala educación, incluso si este no puede darse cuenta. El hecho de que sea incurable, y de que no pueda comprender adecuadamente el diagnóstico, no significa que no esté enfermo. La respuesta de Aristóteles a la pregunta de Sócrates no puede darse a cada persona, como hemos visto, pero es una respuesta *para* cada persona. ¿Dónde cabe situar exactamente esa idea? ¿Qué se dice, en concreto, del hombre malo? No solo estamos afirman-

do que este nos parece un estorbo peligroso (si es que lo hacemos), o que sea una rareza estadística (si es que lo es). Lo que decimos es que carece de determinadas cualidades características de los seres humanos que son necesarias para vivir una vida típicamente humana. Pero tenemos que decir más, si queremos dejar claro un aspecto esencial de la filosofía de Aristóteles y de cualquier otra que se le parezca: que es el bienestar de *este hombre* y sus intereses lo que está en juego. Tenemos que decir que este hombre está equivocado respecto a sus intereses y que, de hecho, esta forma de actuar es el síntoma principal de su error.

La idea de que las personas pueden tener «intereses auténticos» diferentes de los que creen que tienen ha generado una vasta bibliografía y un grado casi equivalente de sospecha. El debate parte, generalmente, del uso que de esta noción hacen los autores hegelianos y, siguiendo a Hegel, los marxistas; las aplicaciones que se han hecho de esta idea han sido esencialmente políticas; y, a la luz de esas aplicaciones, las sospechas tienen mucho sentido, dado que la apelación a los intereses auténticos de las personas suele usarse como una razón para coaccionarlas en contra de sus intereses «aparentes» (es decir, tal y como ellas los perciben). Algunas de estas sospechas y críticas, sin embargo, se dirigen erróneamente a la propia noción de intereses auténticos. Incluso si un curso de acción fuera acorde a los intereses auténticos de alguien, el hecho de que este no lo perciba así no significa que, asumiendo que no se le puede persuadir, hará falta coaccionarlo si se quieren defender sus intereses auténticos. Pero, en esas circunstancias, alguna otra justificación será necesaria para que podamos defender sus intereses auténticos. Puede ir a favor de los intereses auténticos de Robinson parar de beber, pero eso no le da a nadie instantáneamente el derecho de impedirle beber. (¿A quién? ¿A nosotros? ¿Al médico? ¿Al Estado?). El mero hecho de que los intereses auténticos no coincidan con los intereses tal y como son percibidos plantea *ya* problemas políticos y éticos.

En la reflexión política, la cuestión de los intereses auténticos plantea aún más problemas, en particular sobre los intereses de clase, en los que aquí no puedo profundizar más. Las líneas básicas del problema, sin embargo, son claras. En primer lugar, la idea de los intereses auténticos no crea ninguna controversia en el caso en que el agente simplemente carezca de una información que, a la luz de sus otras preferencias y actitudes existenciales, podría modificar sus deseos. Alguien piensa que va en la línea de sus intereses beberse esta poción porque cree que es una medicina preparada por su amable farmacéutico, pero si se trata,

en cambio, de cianuro, entonces está equivocado respecto de sus intereses. Lo mismo se aplica a las confusiones en el razonamiento deliberativo, aunque en este caso hay implicadas ciertas cuestiones apremiantes relativas a lo que cuenta como constricción puramente racional sobre el razonamiento deliberativo. Así, muchos filósofos creen que es irracional[11] preferir una satisfacción más inmediata a una satisfacción ulterior solo porque se dé antes, aunque admiten que las diferencias en el grado de certeza afectan a la cuestión en la práctica. Otros consideran evidente que la «proximidad» de las satisfacciones, por decirlo con palabras de Bentham, es en sí misma una dimensión del razonamiento práctico. Alcanzar una conclusión sobre este asunto resulta ciertamente relevante para la cuestión de qué cuenta como un error en la racionalidad del interés propio.

Las cuestiones más significativas sobre los intereses auténticos se plantean cuando el error en el agente va más allá de la falta de información o de la mera racionalidad (sean cuales sean sus límites) y afecta a los deseos y las motivaciones desde los que este delibera; o cuando el error del agente consiste en que no crea algo que racionalmente habría de creer. Un caso paradigmático de esto es el del adolescente desesperado que intenta suicidarse (esto es, que trata realmente de suicidarse, no que comete una tentativa de suicidio). Susan, que acaba de intentar suicidarse, no cree que las cosas vayan a ser diferentes en el transcurso de tres meses, o si lo cree no le importa; no quiere seguir existiendo porque las cosas vayan a mejorar dentro de tres meses. Si creemos que dentro de tres meses todo será diferente y tomamos las medidas necesarias para mantener con vida a Susan, entonces parece que actuamos en beneficio de sus intereses auténticos; cosa que, si estamos en lo cierto, ella reconocerá, pongamos por caso, dentro de seis meses. Esos intereses no están representados en sus motivaciones actuales de una manera más profunda que la que se ha examinado antes. La falta de deseo de vivir en Susan, su incredulidad con respecto a un futuro mejor, forman parte de la condición de la que se curará en tres meses. Su incapacidad para ver cuáles son sus intereses es un síntoma.

---

[11] Por ejemplo, John Rawls, *A Theory of Justice* (Cambridge, Mass., Harvard University Press, 1972, 1.ª ed.) [*Teoría de la justicia*, trad. de M. Dolores González (México, Fondo de Cultura Económica, 1995)], sec. 45; véanse también secs. 63 y 45. Para la posición contraria, véase Derek Parfit, *Reasons and Persons* (Oxford, Clarendon Press, 1984) [*Razones y personas*, trad. de M. Rodríguez González (Madrid, A. Machado Libros, 2004)].

Pero no podemos limitarnos a decir que un cambio iría en beneficio de los intereses auténticos de alguien porque, como resultado del cambio, esa persona reconocería que este ha beneficiado a sus intereses. Quizá, si una secta religiosa nos lavara el cerebro, identificaríamos completamente nuestros intereses con los de esa secta. Nuestra condición de creyentes, fruto del lavado de cerebro, podría hacer que tuviéramos mucho que decir acerca del incremento en iluminación y comprensión que hemos experimentado a partir de nuestra ceguera anterior, pero ello no establecería el valor del lavado de cerebro. Estas dificultades se plantean con cualquier proceso psicológico que tienda a generar una creencia. Una reacción ante tales dificultades es tirar la toalla y considerar que la noción de intereses auténticos es inevitablemente subjetiva o, quizá, ideológica. Con todo, subsiste un verdadero problema por el mero hecho de que hay ciertas restricciones relativas a lo que podemos considerar aceptablemente como la mejora de la situación de una persona a resultas del cambio, en oposición a que las cosas sean en general mejores o a que salgamos beneficiados nosotros mismos. «Estaría mejor muerto» puede decirse de alguien por muchas razones sospechosas: la más sospechosa es que nos beneficiaríamos de que esa persona estuviera muerta.

Si hay un punto de apoyo más firme para la noción de interés auténtico (y parece que incluso las posiciones más escépticas al respecto requieren restricciones suplementarias), este debe ir en la dirección de excluir los cambios que se validan a sí mismos, como el caso del lavado de cerebro. Una propuesta natural es la siguiente: si un agente no reconoce ahora que determinado cambio iría en beneficio de sus intereses pero, como resultado de ese cambio, llegara a reconocerlo, esto mostraría que el cambio iba realmente en beneficio de sus intereses solo con la condición de que la alteración de su perspectiva se explique por una *incapacidad general* que padecía en su estado original, y que ha desaparecido o se ha reducido gracias al cambio. «Incapacidad general» es una expresión vaga, pero transmite dos ideas relevantes. Una es que la supuesta incapacidad del agente para identificar sus intereses auténticos antes del cambio no dependa sencillamente de que esté hecha a medida del contenido del cambio recomendado, sino que tenga implicaciones más generales, lo que no ocurría en el caso de la supuesta incapacidad para reconocer los méritos de la secta religiosa. La segunda idea es que lo que esté en cuestión sea, de hecho, una *incapacidad*. No se trata solo de que no reconozca ciertas cosas que después del cambio reconocerá,

sino de que la capacidad de reconocer tales cosas en circunstancias culturales de ese tipo es de esperar en los seres humanos, como parte de su funcionamiento eficaz. Es este último elemento, la concepción normativa del funcionamiento humano, el que nos permite usar los términos «curación» y «síntoma» en la descripción del intento de suicidio.

Si de todas formas vamos a tener que introducir esas nociones, ¿por qué no hacerlo mucho antes? ¿Por qué no decir sencillamente que un cambio es favorable a los intereses auténticos de alguien, si el resultado del cambio lo acercaría al funcionamiento humano normal? La respuesta es que no todo lo que es beneficioso para los intereses auténticos de una persona es necesario para su funcionamiento humano o algo que esta *necesite*. Lo que necesita son capacidades, incluidos los patrones de motivación básicos, para intentar conseguir algunas de las cosas que favorecen sus intereses. Si esto no es puramente ideológico, la idea de intereses auténticos necesita estar acompañada de una teoría del error, una explicación sustantiva de la manera en la que las personas yerran en el reconocimiento de sus propios intereses auténticos[12].

El propio Aristóteles sostuvo una teoría muy radical sobre la teleología general: cada tipo de cosa tenía una forma ideal de funcionamiento que encajaba con la de las demás cosas. Creía que todas las excelencias de carácter tenían que encajar en un yo armonioso. Además, estaba comprometido con la idea de que los desarrollos más elevados de la naturaleza humana, que identificaba con la indagación intelectual, encajarían a la perfección con la vida más cotidiana de la virtud cívica, incluso aunque representaran el florecimiento de poderes bastante diferentes, de la razón teórica más que de la práctica. Pero no tuvo mucho éxito en su empeño por demostrar esto. Asimismo, al margen de los ricos recursos teleológicos de su teoría, no hizo gran cosa para propor-

---

[12] Hay una analogía entre la caracterización propuesta de los intereses auténticos y determinadas consideraciones de la teoría de los intereses puramente cognitivos. En muchas circunstancias de error potencial, las investigaciones conducidas en tales circunstancias mejorarán nuestra información. Una consecuencia es que podemos aprender a no ser engañados. Hay otras circunstancias de error, sin embargo, tales como soñar, en las que esto no es posible (y por ello han contribuido a suscitar el escepticismo filosófico). La explicación de lo que está mal en tales circunstancias también da cuenta de por qué esa explicación no puede aplicarse en esas circunstancias. Véase mi libro *Descartes: The Project of Pure Enquiry* (Atlantic Highlands, N. J., Humanities Press, 1978) [*Descartes: el proyecto de una investigación pura,* trad. de J. A. Coll (Madrid, Cátedra, 1996)], cap. 2 y apéndice 3.

cionar la teoría del error que requiere la noción de intereses auténticos. Describe diversos tipos de malas personas y sus descripciones son más realistas que las de Platón en la *República* —quien (en esta obra, pero no en todas) cae en la tentación moralista de representar a la mala persona como un adicto compulsivo, un desecho poco envidiable—. Aristóteles entiende, en cambio, que alguien puede dejar mucho que desear desde el punto de vista ético sin ser en absoluto un desecho; en particular, puede ser capaz de usar la razón de manera eficiente para tratar de conseguir lo que supone ventajoso para sí. Aristóteles explica la condición de esa persona afirmando que su educación fue deficiente, lo que le llevó a adquirir el hábito de buscar placeres inadecuados. Pero en el universo teleológico de Aristóteles todo ser humano (o, al menos, todo varón sin taras que no sea esclavo por naturaleza) tiene un tipo de conato interno que le impulsa a llevar una vida virtuosa o, por lo menos, de virtud cívica. Sin embargo, Aristóteles no dice lo suficiente acerca de cómo esta inclinación puede frustrarse por una educación deficiente ni aclara por qué, tras esa educación, todavía resulta beneficioso para los intereses auténticos de este hombre ser quien no es.

Si Aristóteles, con sus fuertes presupuestos relativos al conato de cada especie natural por alcanzar su perfección, no pudo establecer esta explicación de manera firme, entonces no cabe esperar que podamos hacerlo nosotros. La biología evolutiva, que constituye nuestra mejor comprensión de los hechos que Aristóteles presentaba en términos de la metafísica teleológica, no puede llegar a mejores resultados en su intento de demostrar que la vida ética es una vida de bienestar para cada persona. Esto no se debe a que proporcione una respuesta única para todos los individuos, sino a que su respuesta es hostil a la vida ética, como la que dice que una estrategia «agresiva» sería beneficiosa para todos y cada uno de los individuos: no podría ser así, puesto que el resultado no constituiría un «estadio evolutivo estable», en términos de John Maynard Smith[13]. Lo importante es que la biología evolutiva no

---

[13] John Maynard Smith, *Evolution and the Theory of Games* (Nueva York, Cambridge University Press, 1982). Michael Ruse, *Sociobiology: Sense or Nonsense?* (Hingham, Mass., Kluwer Boston, 1979) [*Sociobiología*, trad. de A. Martín Santos (Madrid, Cátedra, 1989)] da buena cuenta de algunos aspectos fundamentales. Para la definición de «eficacia biológica» *[fitness]* usada en este texto y su importancia, véase Richard M. Burian, «Adaptation», en Majorie Greene (ed.), *Dimensions of Darwinism* (Nueva York, Cambridge University Press, 1983). Para otras consideraciones sobre la biología evolutiva y la ética, véase «Evolution, Ethics and the Representation Problem», en

se ocupa directamente del bienestar del individuo, sino de la eficacia biológica, que es la probabilidad de que ese individuo deje descendencia. Lo máximo que la sociobiología podría aportar a la ética va en una dirección distinta, si llegara a ser capaz de mostrar que determinadas instituciones o patrones conductuales no son opciones realistas para las sociedades humanas. Este sería un logro importante, pero antes hará falta que la sociobiología sepa interpretar la historia de la cultura humana mucho mejor de lo que es capaz de hacerlo hoy en día.

Si alguna ciencia puede llegar a conclusiones que sean válidas *para* cada una de las personas, como dije antes, será alguna rama de la psicología. Existen teorías, especialmente las de tipo psicoanalítico, de las que se espera que den apoyo a que alguna concepción ética sea parte necesaria de la felicidad humana. En algunos casos, las teorías producen esta impresión porque son ya pensamiento ético[14]. No son peores por ello, como cauces para ayudar a los individuos, y probablemente serán mejores, pero esto las descalifica como explicación independiente del bienestar y, por tanto, fundamento de la vida ética. Quizá sea poco realista suponer que puede haber alguna disciplina psicológica capaz de hacer esto. Sería insensato tratar de determinar *a priori* y en unas pocas páginas si puede haber una teoría semejante. Tendría que ser, a la vez, independiente de las concepciones éticas adoptadas, estar estrechamente relacionada con los complejos aspectos de la personalidad humana que están implicados en la vida ética, llegar a resultados definitivos y —por supuesto— ser favorable a las consideraciones éticas de una forma u otra. Este último punto es indispensable «por supuesto», no solo a causa de la tediosa razón de que solamente entonces la teoría sería capaz de proporcionar un fundamento para las consideraciones éticas, sino porque, en caso de no ser favorable a estas, tendría una relación completamente distinta con la práctica. Necesitamos vivir en sociedad —y esta es, sin duda, una necesidad interna, no solo una necesidad tecnológica— y, para ello, unas u otras consideraciones éticas tienen que encarnarse en las vidas de un gran número de personas. Así pues, una teoría psicológica que mostrara que ningún conjunto de conside-

---

D. S. Bendall (ed.), *Evolution from Molecules to Men* (Nueva York, Cambridge University Press, 1983).

[14] Para esta dimensión del pensamiento de Freud, véase Philip Rieff, *Freud: The Mind of the Moralist,* 3.ª ed. (Chicago, University of Chicago Press, 1979) [*Freud: la mente de un moralista,* trad. de A. Cucurullo (Madrid, Paidós, 1996)].

raciones éticas puede hacernos felices no nos diría cómo hemos de vivir, sino que predeciría, más bien, que no podemos vivir felices.

Cualquier psicología del carácter adecuada incluirá, presumiblemente, de una forma científicamente presentable, la verdad de que muchas personas son detestables porque son infelices y viceversa: cuando su infelicidad no sea algo definible en términos éticos, sino que sea sencillamente una infelicidad básica como la miseria, la rabia, la soledad o la desesperación. Este es un hecho bien conocido e impactante, pero no es el único entre una serie de hechos igualmente cotidianos. Algunas de las personas que no son detestables y se esfuerzan en ser generosas y en amoldarse a los intereses de los demás son desgraciadas, desde el punto de vista de su estado ético. Pueden ser las víctimas de una asertividad reprimida, de la que han sido conscientes, pero que ahora no reconocen y que menos aún pueden superar o corregir. Tenemos también la figura, menos frecuente de lo que suponía Calicles, pero real, del individuo que es lo suficientemente detestable sin ser por ello infeliz, sino peligrosamente floreciente si nos dejamos guiar por cualquier criterio etológico que se atenga a la apariencia externa, como la mirada brillante y el pelo lustroso. Para quienes quieren basar la vida ética en la salud psicológica es desconcertante que puedan existir personas así. Pero es importante saber en qué medida su existencia —la posibilidad de su existencia, de hecho— es un fenómeno cultural. Parecen más brillantes y refinados en la distancia. Algunos nobles del Renacimiento desempeñan ese rol con más estilo que los sórdidos jefes fascistas, los gánsteres o los magnates, que parecen ser sus principales equivalentes contemporáneos, aunque sean imaginarios. Tal vez nos autoengañamos con respecto al pasado. O, quizá, es un logro del mundo moderno haber hecho imposible la perpetuación de esos personajes, porque ha convertido el mal, como otras cosas, en una empresa colectiva, un proceso que lo hace más poderoso aunque más banal[15].

Por encima de todo, en lo referente a esta panoplia de percepciones ingenuas, se sitúa el problema de los otros bienes. Determinado hombre es detestable y mezquino, pero tiene éxito y disfruta de ciertos pla-

---

[15] Una exposición destacada de este fenómeno es la de Hannah Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil* (Nueva York, 1963; Penguin Books, 1977) [*Eichmann en Jerusalén: un estudio sobre la banalidad del mal*, trad. de C. Ribalta (Barcelona, Debolsillo, 2006)]. Con el bien, el efecto es el opuesto.

ceres, y si fuera menos detestable no tendría éxito, pero no estaría menos ansioso, porque estaría frustrado... Aceptar como valiosas solo las disposiciones éticas —es decir, orientar el concepto de salud psicológica hacia la renuncia de otros valores— supondría regresar al ascetismo socrático y requeriría una reconstrucción consecuente del yo. Exigiría también una política utópica de renuncia por parte de todo el mundo; o si no habría que admitir que la virtud como pureza de corazón, aunque sería el único bien, solo podría ser el logro de una minoría, y se precisaría, a su vez, la puesta en marcha de otra política para construir la relación de esa virtud con una sociedad no regenerada.

Esos problemas adquieren una relevancia especial, tanto para el individuo como para el resto de la sociedad, cuando los «otros bienes» son de tipo creativo y cultural. Ya me he referido a que esta es una de las tensiones de la concepción aristotélica de la naturaleza humana al considerar que esos logros están armónicamente integrados con el conjunto de las virtudes cívicas cotidianas. Pensar que esta armonía es imposible no es una idea específicamente moderna; de hecho, Platón adoptó una concepción mucho más pesimista y consecuentemente habría deseado el destierro de las artes de la república virtuosa o, al menos, su domesticación. Pero las concepciones modernas de las artes y las ciencias y de la psicología de su creación solo pueden agudizar, desde una perspectiva ética, los problemas de la herida y el arco*, de la infelicidad y la fealdad susceptibles de intervenir en la actividad creativa, la cual suele estar determinada por un desequilibrio, una hipertrofia de determinados poderes y sensibilidades.

Este es también un problema para todo programa que quiera conectar la vida ética con la salud psicológica por medio de las nociones de integración o de reducción de los conflictos. Esos objetivos psicológicos no pueden tener una importancia ética por sí mismos, a no ser que estén ya definidos con miras a ello; la mejor manera de integrar a ciertas personas sería hacerlas más despiadadas. Pero aparte de esto, y

---

* Williams utiliza aquí el título del libro de Edmund Wilson *The Wound and the Bow: Seven Studies in Literature* (Cambridge, Mass., Riverside Press, 1941) [*La herida y el arco*, trad. de M. Uribe (México, FCE, 1983)] para referirse a la relación entre la enfermedad y el don artístico a partir del capítulo en el que Wilson trata el episodio de Filoctetes durante la guerra de Troya. Filoctetes pudo sobrevivir nueve años a la soledad y el sufrimiento gracias al poder creativo de su mente. Agradezco a M. Garrido esta información. *[N. del T.]*.

dejando también de lado los conflictos creativos que nos inducen a preguntarnos más bien en qué medida la reducción de los conflictos puede ser psicológicamente deseable, se plantea otra cuestión: la de en qué medida y en qué circunstancias la eliminación del conflicto puede ser *éticamente* deseable. El conflicto, en concreto el conflicto ético, puede ser la respuesta apropiada a ciertos tipos de situación. La eliminación de estas situaciones no será cuestión solo (o no principalmente) de reformar la psique, sino también de cambiar la sociedad.

Al considerar las dificultades de la psicología para realizar una contribución sustantiva a los fundamentos de la ética, es importante que nos preguntemos por lo lejos que podríamos ir sin ella. Y podemos ir bastante lejos. La formación de disposiciones éticas es un proceso natural en los seres humanos. Esto no significa que sean un proceso espontáneo que prescinda por completo de la educación: en este sentido, casi nada en los seres humanos es «natural», ni siquiera el uso del lenguaje (porque, aunque la capacidad de aprender una lengua es en sí misma innata, y muy probablemente específica[16], ningún niño aprenderá a hablar a menos que esté expuesto a una lengua particular, que es, por supuesto, un producto cultural). Esto no significa que la vida ética no comporte convenciones: para los seres humanos es natural vivir entre convenciones. No es *más* natural, como asumía Trasímaco, vivir fuera de las consideraciones éticas. Además, nosotros mismos (la mayoría de nosotros) nos identificamos con ciertas consideraciones éticas y tenemos una concepción del bienestar humano que da cabida a esas consideraciones. Deseamos, consecuentemente, criar a nuestros hijos de forma que compartan algunas de esas concepciones éticas u otras ideas culturales y consideramos que el proceso es bueno no solo para nosotros sino también para ellos, tanto porque es parte de nuestra concepción de su bienestar como también porque, incluso en concepciones más limitadas de la felicidad o de la satisfacción, no tenemos ninguna razón para creer que serán más felices fuera de las instituciones éticas de

---

[16] Véase el trabajo de Noam Chomsky; en particular, *Language and Mind* (Nueva York, Harcourt Brace Jovanovich, 1972) [*El lenguaje y el entendimiento*, trad. de J. Ferraté (Barcelona, Seix Barral, 1977)]. Para una caracterización general, véase John Lyons, *Noam Chomsky* (Londres, 1970; Nueva York, Penguin Books, 1978) [*Noam Chomsky*, trad. de J. Romà Vergés (Barcelona, Grijalbo, 1974)].

la sociedad. Aunque sepamos que hay algunas personas más felices, según criterios mínimos, fuera de esas instituciones, también sabemos que casi nunca llegan a esa situación siendo educados como forajidos. A resultas de todo eso, tenemos más razones a favor de criar a nuestros hijos dentro del mundo ético en el que habitamos que para hacer lo contrario, y si tenemos éxito, ellos verán el mundo desde la misma perspectiva.

Si se acepta este desplazamiento de la pregunta de Sócrates implícita en lo que admitía Aristóteles —que nadie puede regenerar su vida solo con contestar a esta pregunta—, podemos responder a ella, en cierta medida, con menos de lo que Aristóteles ofrecía. Una vez desplazada, se convierte en la pregunta de cómo hemos de vivir *nosotros* y, a cierto nivel, podemos darle una respuesta ética en función de la vida ética que llevamos, incluso si no podemos afirmar, como hacía Aristóteles, que disponemos de una respuesta teleológica para cada persona que favorezca la vida ética. A este nivel, se trataría de saber si la sociedad habría de ser reproducida éticamente, y a esta pregunta solo podemos dar una respuesta desde dentro de la sociedad.

A este nivel. El argumento presenta solo la elección entre una vida ética y una vida no ética. Pero la vida ética no es una cosa dada de manera unitaria. Comporta diversas posibilidades de educación, de decisión social o, incluso, de regeneración personal. La vida ética que llevamos tiene sus diversidades, incoherencias e instrumentos de autocrítica. Todo esto está más presente en nuestra sociedad moderna que, quizá, en cualquier sociedad que haya existido, y este es un hecho de la mayor importancia, que cambia la función de la filosofía moral de una manera que no podría haber imaginado ningún autor antiguo. Tras considerar, en el capítulo siguiente, un intento muy diferente y más moderno de hallar unos fundamentos para la empresa ética en su conjunto, pasaré a ocuparme de las cuestiones que plantean esa diversidad de posibilidades para la vida ética: algunas conciernen a culturas diferentes, otras a la nuestra. Veremos cómo la filosofía moral puede ayudarnos a entenderlas y a fundar la autocrítica.

En primer lugar, sin embargo, cabe considerar la distancia que separa la concepción de Aristóteles de los estilos de pensamiento ético aceptables hoy en día. En muchos aspectos sustantivos, como he dicho, ninguna discusión moderna puede compartir la perspectiva de un autor antiguo. Pero, ¿en qué medida se transmite esto al contorno lógico de toda esta empresa? He dicho que, si se ha de satisfacer la exigencia

socrática en su forma original y si la vida ética ha de estar justificada para cada persona, el valor ético tiene que residir en algún estado del yo. Ocurriría lo mismo si, según mi formulación del objetivo de Aristóteles, pudiera darse esta justificación solo *para* cada persona. Pero, si renunciamos a ese objetivo también, ¿se hallará el valor ético todavía en los estados de la persona? Suele pensarse que esta manera en la que los griegos abordaron el pensamiento ético es lo que distingue la ética griega de las concepciones modernas. Además, esa idea se asocia a una versión más profunda de una crítica que ya he mencionado antes, según la cual el pensamiento ético griego sería incorregiblemente egoísta.

A favor de este juicio, puede aducirse el siguiente argumento: la persona virtuosa en el sentido aristotélico desea, bastante a menudo, hacer diversas cosas virtuosas. Pero cualquier cosa motivada por el deseo se orienta al placer, y la búsqueda del placer es egoísta. La única motivación opuesta a esta es el sentido de obligación. La motivación ética comporta una oposición al egoísmo, de modo que la ética tiene que implicar la obligación y ser ajena a los deseos de vivir una vida de bienestar, como supuso Aristóteles. (Esta manera de ver las cosas constituye una versión especialmente sintética y tosca del punto de vista de la moral). Casi todos los presupuestos de este argumento son erróneos. Es falso, y también incoherente, suponer que todo deseo aspira al placer. Y esto sería falso incluso si la satisfacción de todo deseo abocara al placer (cosa que, en todo caso, no es así). Además, si fuera verdad que todo deseo aspira al placer no podríamos admitir el presupuesto de sentido común según el cual hay una oposición entre las motivaciones éticas y la búsqueda del placer. Las motivaciones éticas tendrían por objeto cierto tipo de placeres. Algunas de ellas producen de hecho placer, y Hume, en consonancia con el pensamiento griego en este punto, creía con agrado que era una marca de la persona virtuosa encontrar placer al realizar acciones generosas y serviciales.

Una razón evidente por la que no todos mis deseos tienen como objetivo mi placer es que algunos de mis deseos son relativos a que existan estados de cosas en las que yo no tengo ningún papel: no se me menciona en la descripción completa de lo que satisfaría ese deseo[17]. Hay deseos que trascienden al yo. Estos no son todos altruistas o bene-

---

[17] He analizado esto con más detenimiento en «Egoism and Altruism», en *Problems of the Self* (Nueva York, Cambridge University Press, 1973) [«El egoísmo y el altruismo», *Problemas del yo*, trad. de J. M. G. Holguera (México, UNAM, 1984; 2.ª ed. rev., 2013)].

volentes, también pueden ser maliciosos o frívolos. Quienes incluyen en sus testamentos cláusulas destinadas a mortificar a sus parientes o a promocionar un proyecto absurdo no suelen creer que estarán allí para disfrutar del resultado; pero, aun así, lo que quieren es que se dé este *resultado* y no solo el placer de pensar sobre él ahora. Por todas estas razones, la línea de demarcación entre lo que me concierne y lo que concierne a los demás no se corresponde de ningún modo con la línea divisoria entre el deseo y la obligación. (De hecho, algunos moralistas admiten esto a su manera inventando una clase de *deberes hacia uno mismo,* obligaciones que conciernen al propio sujeto. Tales obligaciones desempeñan una serie de funciones en esta economía: una de las cuales es estimular la inversión a largo plazo contra el consumo inmediato, otra solo supone un blanqueamiento de la moneda del deseo).

Incluso si nos deshacemos de esas concepciones erróneas del deseo y el placer, seguiremos estando sujetos a la acusación de egoísmo. Las disposiciones éticas son disposiciones a querer ciertas cosas, a reaccionar de determinada manera ante los demás y sus acciones, a usar nociones tales como la de obligación, a respaldar ciertos resultados como justos, etc. Probablemente, el agente será parte interesada en las relaciones en cuestión y, por supuesto, es el agente quien reflexiona y decide cómo va a actuar. No es necesario que ninguna de esas concepciones (incluyendo lo que quiere el agente) incorpore una implicación perjudicial para el yo del agente en su contenido; nada de esto implica, en sí mismo, tipo alguno de egoísmo. Pero la pregunta de Sócrates introduce otra idea: implica que el agente *piense* en esas mismas disposiciones y las refiera a una vida de bienestar. Incluso si las disposiciones no se caracterizan por orientarse hacia el yo, aun así es siempre su propio bienestar lo que el agente que se entrega a la reflexión socrática tendrá en cuenta. El egoísmo parece resurgir.

La respuesta a este problema reside en un punto fundamental de la concepción aristotélica, según la cual las disposiciones éticas esenciales residen en el contenido del yo. Llegado el momento de la reflexión madura, yo soy lo que he llegado a ser, y mi reflexión, aunque se refiera a mis disposiciones, tiene que expresar estas disposiciones al mismo tiempo. Pienso en la ética y en otro tipo de bienes *desde* un punto de vista ético que ya he adquirido y que forma parte de lo que soy. Al pensar en bienes éticos y de otro tipo, el agente piensa según un punto de vista que ya sitúa esos bienes en relación los unos con los otros y otorga un significado especial a los bienes éticos. Visto desde fuera, este punto

de vista pertenece a alguien en quien las disposiciones éticas adquiridas están más profundamente arraigadas que otras voliciones y preferencias.

La diferencia entre el punto de vista interno, la perspectiva desde las propias disposiciones y el punto de vista externo a esas disposiciones muestra explícitamente que no es totalmente cierto que los valores éticos descansen en las disposiciones del yo, y no obstante, en otros sentidos, sí es verdad. No lo es desde el punto de vista constituido por las disposiciones éticas —la perspectiva interna— que las únicas cosas de valor sean las disposiciones de las personas; menos aún, que solo las disposiciones del agente tengan valor. El bienestar de los demás, las exigencias de la justicia, entre otras cosas, tienen valor. Si adoptamos la otra perspectiva, no obstante, y miramos las disposiciones de las personas desde fuera, podemos preguntarnos: «¿Qué se precisa en el mundo para que exista el punto de vista ético?». La respuesta solo puede ser: «Las disposiciones de las personas». En un determinado sentido estas son el soporte último del valor ético, lo que tiene un significado tanto práctico como metafísico. La preservación del valor ético radica en la reproducción de disposiciones éticas.

El agente tiene acceso al punto de vista externo a sus disposiciones, pero, si en el curso de la reflexión trata de abstraerse totalmente de esas disposiciones y de pensar en sí mismo y en el mundo como si no las tuviera, entonces no debería sorprenderse si no puede adquirir una imagen adecuada del valor de nada, incluidas sus propias disposiciones. No lo consigue precisamente porque esas disposiciones forman parte del contenido de su yo real. Además, si va a llevar a cabo una reflexión que lo distancie de sus propias disposiciones, es relevante saber si, en la concepción de las cosas que adopta desde fuera, interviene un elemento que entre en conflicto con la concepción de las cosas que adopta desde dentro[18]. Para Aristóteles, el agente virtuoso no hallaría ningún conflicto de este tipo. Podría entender que las disposiciones de las que extrae su concepción ética del mundo representan un desarrollo normal y completo de las potencialidades humanas. Esto era así *sin paliativos* en el sentido, al menos (Aristóteles, sin duda, iba más allá), de que la mejor teoría posible de la humanidad y de su lugar en el mundo produciría este resultado. Ade-

[18] La importancia de este requisito —que un punto de vista externo desde el que puedo comprender mis disposiciones no habría de distanciarme de ellas— reaparece (en relación con un sentido diferente de «comprender») en el capítulo 6.

más, esta perfección podría desplegarse armónicamente, de modo que el desarrollo de esas capacidades éticas encajara con otras formas de excelencia humana. La teoría de Aristóteles implica que si el agente reflexiona, incluso desde fuera, sobre todas sus necesidades y capacidades, no estará nunca en conflicto con sus disposiciones éticas.

Aquí reaparecen las múltiples dudas modernas que debilitaban esta teoría. Nuestra comprensión actual no nos da razón alguna para esperar que las disposiciones éticas puedan, en último término, armonizarse plenamente con las demás aspiraciones culturales y personales que pueden pretender representar igual de bien el desarrollo humano. Aunque dejemos la puerta abierta a una psicología que pueda avanzar en la dirección aristotélica, es difícil creer que una caracterización de la naturaleza humana —si no es ya una teoría ética— llegue a definir adecuadamente un tipo de vida ética que se contraponga a las demás. Aristóteles entendió que determinado tipo de vida ética, cultural e incluso política, era la culminación armoniosa de las potencialidades humanas adquirida a partir de una comprensión absoluta de la naturaleza humana. Nosotros no tenemos ninguna razón para creer eso. Por consiguiente, se abre un abismo potencial entre la perspectiva del agente y el punto de vista externo. Comprendemos —y, lo que es más importante, el agente puede llegar a comprender— que la perspectiva del agente es solo una entre las muchas que son igualmente compatibles con la naturaleza humana, todas sujetas a diversos conflictos en el interior de sí mismas y con otras metas culturales. El abismo abierto entre ambas perspectivas hace que la tesis que expresé diciendo que las disposiciones del agente son el «soporte último» del valor ético adopte un tono más escéptico. Nunca más parecerá suficiente.

Creo que esta tesis es verdadera y que en esencia la descripción del yo ético que hemos heredado de los antiguos es correcta. Pero, al mismo tiempo, tenemos que admitir que los presupuestos aristotélicos que conciliaban la perspectiva del agente y la concepción externa se han derrumbado. Nadie ha encontrado aún una manera satisfactoria de prescindir de esos presupuestos. Este es el estado de cosas sobre el que se articulará el argumento de este libro, y volveré en diversas ocasiones a las relaciones entre los puntos de vista interno y externo. Antes, no obstante, examinaré otra forma de proceder que intenta ir de abajo arriba, que trata de hallar un punto arquimédico sin recurrir a los presupuestos aristotélicos.

Capítulo 4

Los fundamentos: la razón práctica

El proyecto del capítulo anterior, que no era otro que el de fundamentar la vida ética en el bienestar, aspiraba a extraer determinadas conclusiones sobre la forma de una vida completa a partir de creencias sustantivas sobre la naturaleza humana. Vimos que para dar coherencia al conjunto de estas creencias hacían falta presupuestos muy fuertes, presupuestos que nosotros no podemos aceptar.

Existe otro proyecto que intenta también partir del fundamento ir de abajo arriba, pero que nos da menos por menos. No nos ofrece la caracterización de una vida completamente desarrollada, sino solo ciertos rasgos estructurales o formales de las relaciones éticas. En lugar de apoyarse en una teleología específica de la naturaleza humana, parte de una concepción muy abstracta de la agencia racional. Aspira también a responder a la pregunta de Sócrates, siquiera de una manera mínima. Da una respuesta válida para todo agente, en tanto que cualquier agente puede plantearse la pregunta. Por ello, sus respuestas son más abstractas y menos específicamente humanas que las aristotélicas. En todo caso, esta clase de argumento está vinculado a principios generales y formales reguladores de las relaciones entre agentes racionales. Esto es lo que le interesa a Kant.

Puede parecer sorprendente decir esto. El nombre de Kant se asocia a un planteamiento en el que, se suele suponer, no puede haber ningún *fundamento* para la moral. Kant pone énfasis en que la moral ha de ser

«autónoma» y en que no puede haber razón alguna para ser moral. Un argumento simple muestra por qué, en el marco kantiano, esto tiene que ser así. Cualquier razón para ser moral ha de ser ya una razón moral o no moral. Si es moral, entonces no puede ser realmente una razón para ser moral, pues uno tendría que estar ya dentro de la moral para aceptarla. Una razón no moral, por otro lado, no puede ser una razón para ser *moral;* la moral exige una pureza de motivo, una intencionalidad básicamente moral (que Kant entendió que era la obligación), y esto se desvanece al aceptar cualquier incentivo no moral. Por lo tanto, no puede haber ninguna razón para ser moral. La moral se presenta a sí misma como una exigencia inmediata, como un imperativo categórico.

Lo que Kant presenta es específicamente *la moral.* Abordaremos con más detalle cuestiones relativas a esta concepción de la vida ética cuando retomemos el asunto en el capítulo 10. La perspectiva de Kant exige que no haya ninguna razón para la moral, si ello significa una motivación o incentivo para ser moral, pero no implica que la moral carezca de fundamentos. Kant pensaba que éramos capaces de comprender por qué la moral ha de presentarse al agente racional en forma de exigencia categórica. Esto sería así porque la agencia racional misma implicaría la aceptación de esa exigencia, que es por lo que Kant describió la moral en términos de leyes que la razón práctica promulga por sí misma[1].

En su extraordinario libro *La fundamentación de la metafísica de las costumbres,* la obra más importante de la filosofía moral desde Aristóteles y una de las más desconcertantes, Kant trata de explicar cómo puede ser esto así. Pero no quiero exponer el argumento explicando directamente a Kant. Esto conllevaría, de suyo, muchos problemas específicos. Trataré su posición más como el punto de llegada que como un itinerario. Comenzaré por un argumento más simple y concreto que el de Kant[2]. El fracaso de este argumento a la hora de hallar un funda-

---

[1] Quizá la posición de Kant puede resumirse mejor diciendo que este nos ofrece una caracterización de la moral y una caracterización de la razón práctica, y sostiene que ambas llegan al mismo resultado.

[2] El argumento que desarrollaré es semejante en muchos aspectos al presentado por Alan Gerwith en *Reason and Morality* (Chicago, University of Chicago Press, 1977). Gerwith dice: «Aunque la importancia de la acción para la filosofía moral ha sido reconocida desde los antiguos griegos, no se ha señalado hasta la fecha que la naturaleza de la acción entra en el contenido y en la justificación del principio supremo de la moral»

mento para la moral nos ayudará a mostrar por qué la concepción kantiana necesita ser tan metafísicamente ambiciosa como es. No creo que el argumento de Kant tenga más éxito que el mío, pero hay que seguirlo durante un largo trecho para descubrir por qué.

¿Hay algo que los agentes racionales quieran necesariamente? Es decir, ¿hay algo que quieran (o querrían si pensaran en ello lo suficiente) solo como parte o precondición de que sean agentes?

Cuando nos disponemos a actuar, las personas necesariamente queremos, en primer lugar, cierto resultado: queremos que el mundo sea de una manera y no de otra. Uno puede querer que se dé un resultado sin querer ser quien produce ese resultado; podríamos preferir que el resultado meramente se materialice. Ahora bien, hay ciertos casos en los que el resultado que queremos solo contará como tal si no lo producimos directamente (como cuando queremos que alguien se enamore de nosotros). Pero, en contraste directo con esa posibilidad, en muchos casos lo que queremos esencialmente no es solo el resultado, sino producir nosotros mismos el resultado. Para decirlo de otro modo (que es complicado, pero aún esconde otras complicaciones), el mismo resultado que una persona quiere incluye la acción que la deliberación presente de esa persona convertirá en su hacer[3].

No solo queremos que el mundo contenga ciertos estados de cosas (es un craso error del consecuencialismo creer que esto es todo lo que queremos). Entre las cosas que básicamente queremos, está actuar de cierta manera. Pero, incluso cuando lo que básicamente queremos es que se den ciertos estados de cosas, y nos haría felices que se materiali-

---

(pág. 26). Aunque la caracterización de Gerwith contiene muchos rasgos originales, creo que subestima la afinidad de su proyecto con el de Kant. Si bien su tratamiento de estos asuntos difiere en ciertos aspectos de los considerados aquí, creo que fracasa por las mismas razones generales.

[3] No basta con decir: el resultado deseado incluye que hayamos hecho algo. Esta fórmula podría aplicarse igual de bien al tipo de caso anterior; por ejemplo, queremos que el resultado incluya no solo que alguien se enamore de nosotros, sino también hacer el amor con esa persona. Para complicaciones ulteriores, y sobre la relevancia de todo esto para el consecuencialismo, véase «A Critique of Utilitarianism», en J. J. C. Smart y Bernard Williams, *Utilitarianism: For and Against* (Nueva York, Cambridge University Press, 1973) [*Utilitarismo. Pro y contra,* trad. de J. Rodríguez Marín (Madrid, Tecnos, 1981)], sec. 2.

zaran, sabemos que no vivimos en un mundo mágico, en el que querer que se dé un resultado basta para que se produzca. Conscientes, por tanto, de que ese estado no se producirá a no ser que actuemos para que se produzca, cuando queremos que se dé un resultado, solemos querer también producirlo. (Con esto guarda una analogía directa el principio que afirma que, cuando queremos la verdad, queremos saber la verdad)[4]. Además, no solo queremos que *resulte* que lo hemos producido; queremos que nuestros pensamientos lo produzcan. Así, las voliciones implicadas en nuestras actividades intencionales resultan ser muy complejas. Como poco, lo que queremos es que el resultado se produzca porque lo queremos, porque creíamos determinadas cosas y porque actuamos como lo hicimos sobre la base de eso que queremos y creemos[5]. Consideraciones semejantes se aplican a la conservación de las cosas que queremos conservar.

Esto se añade, pues, a lo siguiente: en diversas ocasiones queremos determinados resultados; normalmente queremos producir esos resultados; usualmente queremos producirlos de un modo que se corresponda con el hecho de que queremos producirlos. Como es bastante obvio, en esas ocasiones no queremos quedar frustrados, por ejemplo, por la intervención de otras personas. Reflexionando sobre todo esto, podemos ver que tenemos una volición general, disposicional, de no quedar frustrados, en particular por la intervención de otras personas. Tenemos una volición general, por decirlo sumariamente, a favor de la libertad. Esto no desdice que a veces queramos perder libertad, quedar frustrados a causa de otras personas o incluso ser coaccionados; pero, en ese caso, no queremos que se nos frustre la obtención de esta *frustración*[6].

No obstante, no basta para disfrutar de esta libertad con que no se nos impida hacer aquello que queremos hacer. Podríamos ser capaces de hacer todo lo que queremos, simplemente porque queremos pocas cosas. Podríamos tener artificialmente voliciones precarias o empobrecidas. Esta consideración muestra que tenemos otra volición general,

---

[4] Véase mi libro *Descartes: el proyecto de una investigación pura,* cap. 1.

[5] Este tipo de estructuras son características del fenómeno de la intencionalidad. Véase, en particular, H. P. Grice, «Meaning», *Philosophical Review,* 66 (1957) [en castellano en L. M. Valdés (ed.), *La búsqueda del significado,* 3.ª ed. (Madrid, Tecnos, 1999), págs. 485-494; trad. del artículo de A. Menasé], y John Searle, *Intentionality* (Nueva York, Cambridge University Press, 1983) [*Intencionalidad,* trad. de E. Ujaldón Benítez (Madrid, Tecnos, 1992)].

[6] Como señala Gerwith, pág. 53.

aunque sea indeterminada: queremos (dicho vagamente) tener una gama adecuada de voliciones.

No se sigue de todo esto que queramos que nuestras elecciones estén tan poco limitadas como sea posible, por nada ni nadie. No queremos que nuestra libertad no tenga límites. Puede parecer que sí se sigue[7], pero aceptarlo sería dejar fuera otra condición vital de la agencia racional. Algunas cosas, sin lugar a dudas, son accesibles a un agente en un momento dado y otras no lo son. Además, que algo sea accesible, y en qué medida, depende de características tanto internas como externas al agente. Este elige, hace planes, etc., en un mundo que tiene una configuración factible en función de dónde esté, de quién sea y de qué pueda llegar a ser. El agente no solo sabe que esto es así (lo que significa que está cuerdo), sino que también sabe, mediante la reflexión, que ello es necesario si pretende ser un agente racional. Asimismo, no es coherente que piense que en un mundo ideal no necesitaría ser un agente racional. La existencia de restricciones con respecto a lo que puede hacer es lo que le exige ser un agente racional, y es también lo que le permiten serlo; más que eso, es también la condición que le permite ser cierta persona en particular, vivir *una* vida. Podemos pensar a veces que estamos tristemente forzados a ser agentes racionales y que en un mundo más feliz no sería necesario serlo, pero esto es una quimera; de hecho, es la quimera por antonomasia.

Condiciones similares se aplican al conocimiento del agente. Al actuar en una situación particular, este tiene que querer que sus planes no fracasen por culpa de su ignorancia o de sus errores. Pero, incluso en ese caso particular, no quiere saberlo todo, ni que su acción no tenga ningún tipo de consecuencias imprevistas. No saberlo todo es, una vez más, la condición que permite tener una vida singular (algunas cosas nos son desconocidas, por ejemplo, porque constituirán nuestro futuro). Si no puedo querer coherentemente saberlo todo, entonces tampoco puedo querer coherentemente no equivocarme nunca. No son lo mismo (la omnisciencia no es lo mismo que la infalibilidad), pero existen muchas conexiones entre ambas cosas. Porque, como enfatizaba siempre Karl Popper, tenemos que cometer errores, y reconocerlos, para ampliar nuestro conocimiento.

---

[7] Hobbes puede que creyera que se seguía, pero es difícil distinguir entre lo que creyó que necesariamente supone querer algo y la concepción siniestra que adoptó de lo que, sin un soberano, implicaba conseguir o mantener algo.

Estas últimas consideraciones tienen que ver con cosas que un agente racional no necesita querer, o en realidad que necesita no querer, para ser un agente racional. Se presupone que se trata de un agente finito, encarnado, históricamente situado —el único tipo de agente que concibo, con la excepción marginal o dudosa de las compañías y otras agencias semejantes, y los únicos que (con las mismas excepciones) podrían ser de interés para la ética—. (Incluso quienes creen en Dios, aunque lo tengan por un agente, no deberían considerarlo de interés para la ética). Supongo que esto no es una sorpresa para nadie. Pero tiene ciertas consecuencias importantes, que abordaremos después.

Así pues, como agentes racionales queremos tener lo que llamé sumariamente libertad, aunque eso no significa libertad sin límites. ¿Nos compromete esto a pensar que nuestra libertad es un bien y que es bueno para nosotros ser libres? Una vía para llegar a esta conclusión consistiría en decir que, cuando un agente quiere que se den diversos resultados concretos, tiene que pensar que esos diversos resultados son buenos. Entonces, estaría forzado a pensar que su libertad era algo bueno, en la medida en que está implicada en el logro de esos resultados[8].

¿Es verdad que, cuando queremos algo y los buscamos expresamente, pensamos que es bueno conseguirlo? Esta es una doctrina tradicional expuesta en el *Menón* de Platón y consagrada en una máxima escolástica: *omne appetitum appetitur sub specie boni,* «todo lo que se busca se busca como algo bueno». Pero no me parece que esto sea cierto. Sin duda, en cualquier comprensión normal del bien se requiere un paso ulterior para pasar de decir que alguien quiere algo o ha decidido tratar de conseguirlo a decir que es bueno, o (más aún) que es bueno que lo tenga. La idea de que algo es bueno transmite la idea, por mínima o vaga que sea, de una perspectiva en la que eso puede ser reconocido como bueno por más de un agente. Un agente que tiene sin más cierto objetivo puede pensar, por supuesto, que su objetivo es bueno, pero no es necesario que lo haga. Lo máximo con lo que podría comprometerse por tener simplemente un objetivo es, al parecer, con que sería bueno *para él* conseguirlo, ¿pero es siquiera esto necesario? Incluso esta tesis modesta implica una perspectiva que va más allá de las voliciones inmediatas del agente, hasta afectar a sus intereses o a su bienestar a más largo plazo. Valorar algo, incluso en relación con nuestros intereses,

[8] Gerwith sigue esta senda.

como cuando pensamos que esto sería mejor «para mí», es siempre ir más allá del mero querer algo. Yo podría llegar a consagrar todo el valor de mi vida a la satisfacción de un deseo, pero, si lo hiciera, no sería solo porque tenía un único deseo. No tener más que un deseo podría significar que mi vida no tiene ningún valor; encontrar todo el valor a un deseo significa que ese deseo es el único que me *importa*[9].

Incluso si abandonamos la doctrina tradicional, de modo que no tenga que considerar todo lo que quiero como bueno, podría aún ser cierto que habría de concebir mi libertad como un bien. «Bueno para mí», propuse, introduce una referencia a mis intereses o a mi bienestar que va más allá de mis propósitos inmediatos, y mi libertad es uno de mis intereses fundamentales. Así, tal vez tenga que considerar que mi libertad es un bien. Pero, en ese caso, no tengo que caer en el pensamiento falaz de que mi libertad constituye un bien, y punto. Esto sería así solo si fuera bueno, y punto, que yo fuese un agente racional, y no hay ninguna razón por la que los demás hayan de admitir tal cosa. En realidad, no está ni siquiera claro que *yo* tenga que admitirlo. Empezamos a abordar ciertas cuestiones más profundas relativas a la concepción que tengo de mi propia existencia.

Todo lo dicho hasta el momento sobre las condiciones y presupuestos básicos de la acción racional parece correcto. El argumento que trata de proporcionar un fundamento para la moral intenta mostrar que, meramente sobre la base de esas condiciones, todo agente está inmerso en un compromiso moral. Todo agente, según este argumento, tiene que pensar del siguiente modo: dado que necesariamente quiero mi libertad más fundamental, tengo que oponerme a los cursos de acción que la eliminarían. Por lo tanto, no puedo concordar con ninguna manera de disponer las cosas en la que los demás tengan el derecho de eliminar mi libertad más fundamental. Así, cuando reflexiono respecto a qué disposición de las cosas necesito a un nivel más fundamental, me doy cuenta de que tengo que afirmar el *derecho* a mi libertad más fundamental. En efecto, tengo que establecer como regla para los demás que respeten mi libertad. Afirmo este derecho solo porque soy un agente racional con objetivos. Pero, si esto es suficiente como base de mi afirmación, será igualmente suficiente como base para la afirmación de

---

[9] No se trata de que un deseo no sea suficiente para dar a alguien una razón para actuar. Ya he dicho que es suficiente (capítulo 1). La verdad es que no toda razón para actuar se fundamenta en una evaluación.

los demás. Si, como supongo, es legítimo y justo que crea que ellos han de respetar mi libertad, entonces tengo que reconocer que es legítimo y justo que ellos crean que yo habría de respetar la suya. Si concluyo a partir de mi necesidad de ser libre que «los demás no deben interferir en mi libertad», tengo que concluir igualmente a partir de su necesidad de ser libres que «yo no debo interferir en la libertad de los demás».

Si esto es así, entonces las necesidades y voliciones fundamentales de cada persona la comprometen con la moral, una moral de derechos y deberes, y quien rechace dar este paso estará en una especie de conflicto pragmático consigo mismo. Comprometido a ser un agente racional, estaría tratando de rechazar los compromisos que ello necesariamente implica. Pero ¿es correcto este argumento? El último paso, en particular —que, si la mera agencia racional es en mi caso el fundamento de un derecho a no padecer la interferencia ajena, entonces tiene que ser igual en el caso de los demás—, es sin duda correcto. Descansa en la versión más modesta y menos discutible de un «principio de universalización», que se pone en práctica sencillamente con un *porque* o *en virtud de*. Si una consideración particular es realmente suficiente para establecer una conclusión en mi caso, entonces es suficiente para establecerla en el caso de cualquiera. Esto tiene que ser así, si este «suficiente» es realmente suficiente. Si la conclusión que nos aboca a la moral no se sigue, tiene que ser a causa de una premisa anterior. Concediendo que son correctas las premisas iniciales relativas a las voliciones y necesidades de un agente racional, el argumento tiene que errar en el punto en el que afirmo por primera vez mi supuesto derecho.

Es útil considerar lo que podría decir el agente al meditar a fondo sobre sus afirmaciones. Podría formularse así:

Tengo ciertos objetivos.
Necesito la libertad para perseguir estos y otros objetivos.
Por lo tanto, necesito la libertad.
Prescribo: no permitamos que los demás interfieran en mi libertad.

Llamemos a quien formula esta reflexión el agente A. Supongamos de momento que sabemos lo que es una «prescripción», y llamemos a esta prescripción de A, *Pa*. Entonces, A también piensa

Pa es razonable,

que significa que *Pa* es razonable en relación con el hecho de que él, A, es un agente racional. A puede, por supuesto, reconocer que otro agente, B, puede albergar los mismos pensamientos. Él sabe, por ejemplo, que

B prescribe: no permitamos que A interfiera en mi libertad,

y, llamando a la prescripción de B, *Pb,* el principio de universalización exigirá que A admita que

Pb es razonable.

Puede parecer que él ha aceptado que la prescripción de B es razonable en el sentido de una afirmación sobre sí mismo. Esto es lo que exige el argumento a favor de la moral. Pero A no ha dado su aprobación a esto. Él solo ha admitido que *Pb* es razonable en el mismo sentido en que lo es *Pa,* y esto solo significa que *Pb* es razonable en relación con el hecho de que B es un agente racional; es decir, que B es tan racional al hacer su prescripción como lo es A al hacer la suya. No significa que B sería racional aceptando *Pa* (o a la inversa), si al aceptarlo se estuviera comprometiendo a no interferir en la libertad de A.

El mismo problema se nos plantea en relación con la discusión siguiente: no se podría obtener nunca el resultado requerido, el acceso al mundo ético, solo desde la consideración del *he de* o *debo* de la agencia racional, el *he de* de la cuestión práctica. Las razones que tiene B para hacer algo no son en sí mismas razones para otra persona para hacer algo. El *haber de* de la razón práctica se conjuga, como cualquier otro, en segunda y en tercera persona, pero esas formas solamente representan mi perspectiva sobre los intereses y cálculos racionales de los demás, la perspectiva de un «si yo fuera tú». Considerando en estos términos lo que B ha de hacer, yo bien podría concluir que él habría de interferir en mi libertad.

Pero ¿se lo puedo «prescribir»?[10]. ¿Qué significa eso? Innegablemente, yo no quiero que interfiera en mi libertad. Pero, ¿genera esto, en sí mismo, alguna prescripción conducente a obligaciones y derechos? El argumento sugiere que, si no prescribo que los demás no deben inter-

[10] Para profundizar en la cuestión de la prescripción, véase mi análisis de la concepción de R. M. Hare en los capítulos 6 y 7.

ferir en mi libertad, estaré lógicamente obligado a admitir que *pueden* hacerlo; cosa que yo no quiero hacer[11]. El núcleo de este argumento es que o bien tengo que darles el derecho a interferir en mi libertad o bien tengo que negarles ese derecho. El argumento incide, en efecto, en que, si soy coherente, tengo que establecer una regla que prescriba que los demás no han de interferir en mi libertad, y esta regla es imprescindible. Pero, por supuesto, en tanto que es una regla general, me exigirá igualmente que no interfiera en la libertad de los demás.

Pero ¿por qué tengo que prescribir una regla? Si me dedico a establecer reglas, sin duda no formularé una en la que ordene a los demás interferir en mi libertad, ni una en la que se lo permita. Pero hay otra posibilidad: que no considere que esa sea mi labor y no establezca ninguna regla ni en un sentido ni en otro. No se me ha de considerar como alguien que expide autorizaciones. Si hay un sistema de reglas, es evidente que, si estas no dicen nada sobre cierta cuestión (al menos si las reglas tienen un alcance lo suficientemente amplio), esto puede interpretarse como una autorización. La ley, como otras agencias soberanas, puede decir algo con solo guardar silencio. Pero, si no hay ley alguna, entonces el silencio no es un silencio significativo o permisivo: es solo silencio. En otro sentido, por supuesto, las personas «podrían» interferir en mi libertad, pero eso solo significa que no haya ninguna ley que se lo prohíba, permita u ordene. Depende de mí y de lo que yo puedo hacer que el hecho de que los demás «podrían» interferir en mi libertad signifique que realmente «pueden». Como dijo el egoísta Max Stirner: «El tigre que me ataca está en su derecho, y yo que lo abato también estoy en mi derecho. No defiendo contra él mi *derecho,* sino *a mí mismo*»[12].

También puedo preguntar por qué, si voy a prescribir tantas cosas, no he de prescribir, siendo mucho más ambicioso, que nadie interfiera en ninguno de mis posibles propósitos particulares. Sin duda, *quiero* conseguir realizar mis proyectos particulares tanto como cualquier otra cosa, y quiero que los demás no interfieran en ellos. De hecho, mi necesidad de una libertad fundamental derivaba de ese tipo de querer. Pero no cabe duda de que el argumento no me permitirá hacer prescripciones sobre todas mis voliciones particulares.

---

[11] Gerwith presenta un argumento de este tipo, pág. 80.

[12] Max Stirner, *Der Einziger und sein Eigentum;* traducido al inglés por S. T. Byington como *The Ego and His Own,* ed. James J. Martin (Sun City, Cal., West World Press, 1982), pág. 128 [*El único y su propiedad;* hay diversas traducciones al castellano].

El argumento depende de una concepción particular del establecimiento de reglas, una concepción que reside en el núcleo de la empresa kantiana. Si yo estuviera en posición de establecer las reglas que quisiera y de imponerlas como instrumento de opresión, podría promulgar una ley que se aviniera con mis intereses y vulnerara los intereses rivales de los demás. Nadie más tendría una razón para obedecer tal ley, excepto la razón que yo le doy. Pero las leyes que estamos considerando en estos argumentos no son de ese tipo, no llevan aparejadas sanciones externas y no responden a desigualdades entre las partes. Son leyes *nocionales*. La cuestión «¿Qué ley podría yo promulgar?» se convierte, así, en «¿Qué ley podría yo promulgar que pueda esperar razonablemente que los demás acepten?». Cuando reflexionamos sobre el hecho de que todo el mundo hace la pregunta desde una idéntica posición de impotencia —pues esas son leyes para un reino en el que el poder no está en cuestión—, vemos que la pregunta podría ser igualmente «¿Qué ley podría yo aceptar?», y así, finalmente, «¿Qué leyes habrían de existir?».

Si esta es la pregunta, hecha con este espíritu, para tal reino, entonces podemos ver por qué su respuesta habría de seguir el principio fundamental de acción de Kant, el imperativo categórico de la moral, que (en su primera formulación)[13] exige: «Obra solo según una máxima por la cual puedas querer que al mismo tiempo se convierta en una ley universal». Pero el problema pasa, de inmediato, a ser: «¿Por qué ha de adoptar uno esa concepción?». ¿Por qué habría yo de pensar en mí mismo como legislador y —dado que no hay distinción— al mismo tiempo como ciudadano de una república gobernada por esas leyes nocionales? Este sigue siendo un problema abrumador, incluso aunque uno se encuentre ya dentro de la vida ética y esté considerando cómo concebirla. Pero el problema es aún más abrumador cuando se exige de cualquier agente racional una manera semejante de ver las cosas. Lo que hace falta es que el argumento nos diga qué hay de particular en los

---

[13] *Fundamentación,* pág. 88 de la traducción al inglés de H. J. Paton, *The Moral Law* (Totowa, N. J., Barnes and Noble Books, 1978) [Ak. IV, 421; hay varias traducciones al castellano]. En conexión con otra formulación del imperativo categórico, Kant afirma: «El ser racional tiene que considerarse siempre como legislador en un reino de los fines posible merced de la libertad de la voluntad, ya sea como miembro, ya sea como jefe» [Ak. IV, 434; en traducción de R. R. Aramayo (Madrid, Alianza, 2002)]. Esta me parece la expresión más iluminadora de su concepción.

agentes racionales para que se les exija que se conciban a sí mismos como ciudadanos abstractos.

Podría pensarse que la pregunta se responde por sí sola porque los agentes racionales en cuanto tales no pueden ser de otra manera y no hay diferencia entre ellos, pero obtener el modelo de este modo sería muy poco convincente. Lo que nos interesa es lo que cualquier persona dada, por poderosa o eficiente que pueda ser, sería razonable que hiciese en tanto que agente racional, que no es lo mismo que lo que sería razonable que hiciese si fuera un agente racional *y nada más*. De hecho, esta equiparación es ininteligible, puesto que no hay modo alguno de ser un agente racional y nada más. Una prueba más sensata consistiría en preguntar qué sería razonable que las personas hicieran si no supieran nada de sí mismas excepto que son agentes racionales; o bien qué habrían de hacer las personas si supieran algo más que eso, pero no conocieran sus propios poderes ni su posición particulares[14]. Este es un test interesante en ciertos casos; en particular, es un test posible sobre la justicia y, para cumplir esta función, puede ser propuesto a quienes se muestran interesados en ella. Pero no es un test convincente sobre lo que sería razonable que alguien hiciera si no tiene ya un interés en la justicia. A menos que alguien esté ya dispuesto a adoptar un punto de vista imparcial o moral, considerará completamente irrazonable la propuesta de que la manera de decidir qué hacer es preguntarse qué reglas establecería si no tuviera ninguna de sus ventajas actuales o no supiera cuáles son.

El proyecto kantiano, para tener alguna esperanza de éxito, tiene que partir de mucho más atrás. Es esencial que se parezca más al proyecto de Kant que el argumento que acabo de bosquejar. El argumento partía de lo que los agentes racionales necesitan y, aunque lo que se dijo al respecto era cierto, no basta para arrastrar a cualquier persona a la esfera de la moral. Kant partía de lo que, en su concepción, *eran* esencialmente los agentes racionales. Pensaba que el agente moral era, en un cierto sentido, un agente racional y nada más, y presentaba como esencial para su concepción de la moral una concepción metafísica particu-

[14] Este es el test que usa John Rawls en su teoría de la justicia, que examino en el capítulo 5. Véase también Amartya Sen, *Informational Analysis of Moral Principles*, en *Rational Action*, ed. T. R. Harrison (Nueva York, Cambridge University Press, 1979).

lar del agente, según la cual el yo de la agencia moral es lo que llamaba el yo «nouménico», que está fuera del tiempo y de la causalidad, y es, así, distinto de la persona concreta, empíricamente determinada, que cada uno de nosotros normalmente considera que es. Según Kant, esta idea trascendental del yo se descubre al reflexionar sobre las exigencias de la libertad; unas exigencias más profundas que cualesquiera otras que hayamos descubierto en el nivel de indagación que hemos llevado a cabo. Kant no creía que pudiéramos comprender plenamente esta concepción, pero pensaba que podríamos reconocer que es posible y que está implicada tanto en la moral como en la acción racional.

La teoría de Kant presenta grandes dificultades e imprecisiones. En primer lugar, Kant pensaba que todas las acciones, excepto las que responden a un principio moral, se explican no solo de manera determinista sino también en términos del hedonismo egoísta[15]. Solo actuando por un principio moral se puede escapar a estar causalmente determinados por el impulso de lograr placer, como los animales. Y a veces lo recalcaba diciendo que solo las acciones que responden a un principio cuentan como ejercicios de la voluntad (que asimilaba a la razón práctica) y son, por tanto, verdaderamente libres. El resto de nuestras acciones, según este planteamiento, serían solo un producto de la causalidad —de la causalidad «ciega», como con poco acierto suele decirse, ya que, como el mismo Kant reconocía, esta causalidad puede a menudo permitir a los agentes, y ciertamente a los animales, ver muy bien adónde van—.

No entraré en la cuestión de en qué medida la propia teoría de Kant puede superar estas dificultades. Pero, sin duda, cualquier teoría que aspire a proporcionar fundamentos necesitará evitarlas. Estamos interesados en la idea de que la libertad racional presupone las consideraciones éticas, lo cual ha de suponer una libertad con la que el escéptico moral, entre otros, está *ya* comprometido. Kant, o quienquiera que argumente como él, puede decir que el deseo de autonomía individual y de racionalidad compromete al escéptico moral con concepciones que

---

[15] Véase, por ejemplo, el Teorema II de la *Crítica de la razón práctica*. La breve referencia que hago a la posición de Kant sobre el determinismo no tiene en cuenta el misterio fundamental de esta posición; a saber, que Kant pensaba que el determinismo intervenía en todos los acontecimientos, aceptaba que las acciones eran acontecimientos, creía en el libre albedrío y condenaba el «miserable subterfugio», como lo llamaba, de querer hacer compatibles el libre albedrío y el determinismo mediante la adscripción de acciones libres a un tipo particular de causa.

se realizan *totalmente* solo en la ley moral, pero es inútil sostener que el escéptico moral tenga que aspirar a un tipo de libertad racional muy diferente de todo lo que se manifiesta en la inteligencia o deliberación práctica no moral. El compromiso del escéptico con la libertad y la racionalidad no puede desvincularse de cosas de las que tiene experiencia, tales como la diferencia entre la decisión consciente y el hecho de encontrarse uno mismo haciendo algo sin pretenderlo. Además, esta no es solo una consideración dialéctica, sobre la influencia que uno puede esperar ejercer sobre el escéptico, es también una cuestión sobre qué concepción de la libertad racional es razonable mantener.

Lo que estamos buscando, pues, es un argumento que, evitando el extravagante equipaje metafísico del yo nouménico, se adentre lo suficiente en el territorio kantiano para retornar con la conclusión esencial de que los intereses más básicos de un agente racional tienen que coincidir con los intereses derivados de una concepción de sí mismo como ciudadano legislador de una república nocional. El argumento podría desplegarse de la siguiente manera: hemos admitido ya que el agente racional está comprometido a ser libre, y hemos dicho algo sobre las condiciones de esa libertad, pero no hemos llegado aún a una comprensión lo suficientemente profunda de en qué tiene que consistir esa libertad. La idea de un agente racional no es solo la idea de tercera persona de una criatura cuyo comportamiento tiene que explicarse en términos de creencias y deseos. Un agente racional actúa *por* razones, y esto va más allá de que actúe según ciertas regularidades o leyes, incluso si estas remiten a creencias y deseos. Si actúa *por* razones, entonces tiene no solo que ser un agente sino también reflexionar sobre sí mismo como agente, y esto implica que se considere a sí mismo como un agente entre otros. Así, se distancia de sus propios deseos e intereses y los considera desde un punto de vista que no es el *de* sus deseos e intereses. Este es el punto de vista de la imparcialidad. Por lo que resulta apropiado para el agente racional, con su aspiración a ser genuinamente libre y racional, concebirse como hacedor de leyes susceptibles de armonizar los intereses de todos los agentes racionales.

Para evaluar este argumento, es importante tener presente que el tipo de libertad racional que introduce se manifiesta, según Kant, no solo en las decisiones de actuar, sino también en la deliberación teorética, en la reflexión sobre la verdad. No es meramente la libertad como agente —el hecho, hablando en términos generales, de que lo que hago dependa de lo que decido— la que conduce a la posición imparcial,

sino mi libertad reflexiva como pensador, y esto se aplica también al caso del pensamiento factual[16]. En ninguno de ambos casos, suponía Kant, estoy causalmente forzado a llegar a una conclusión: puedo distanciarme de mis pensamientos y experiencias, y lo que de otro modo solo hubiera sido una causa se convierte en *una consideración para mí*. En el caso de que llegue a una creencia mediante la reflexión, el tipo de cosa que se transformará de este modo será una evidencia, o lo que yo considero como tal; podría ser, por ejemplo, una percepción. En el caso de la deliberación práctica, probablemente será un deseo, un deseo que tomo en consideración para decidir qué hacer. Distanciándome de la evidencia o de mis deseos, de tal modo que se tornan consideraciones a la luz de las cuales llego a una conclusión, ejerzo en ambos casos mi libertad racional. Cuando, en el caso práctico, adopto el punto de vista externo a mis deseos y proyectos, puedo suscribir mis deseos originales, así como en el caso factual puedo suscribir mi disposición inicial a creer. En estas condiciones, mi deseo original puede ser, a fin de cuentas, mi motivo para la acción (aunque alguien que use esta imagen puede señalar que, en ciertas ocasiones, lo que finalmente hago no estará motivado por ninguno de los deseos que inicialmente tenía, sino que es el producto radical de mi reflexión)[17].

El hecho de que Kant pretenda aplicar su caracterización de la libertad racional tanto a la deliberación factual como a la práctica desvela el error del argumento kantiano. Lo que dice sobre la reflexión se aplica de hecho a la deliberación factual, pero esto se debe a que la deliberación factual no es esencialmente de primera persona. No puede aplicarse a la deliberación práctica ni consigue imponerle a esta la imparcialidad necesaria, porque la deliberación práctica es radicalmente

---

[16] David Wiggins, «Towards a Credible Form of Libertarianism», en Ted Honderich (ed.), *Essays on Freedom of Action* (Boston, Routledge and Kegan Paul, 1978), presenta argumentos interesantes contra la supuesta simetría y contra la tesis kantiana de que el pensamiento no puede ser racional si está causado, una tesis que, en una versión vulgarizada, forma parte del argumentario de los defensores del libertarismo. Sin embargo, si la verdad es que el problema de la libertad genera una asimetría entre el pensamiento factual y la deliberación práctica, esto no contribuirá a restablecer el tipo de argumento que estoy examinando: su intento de fundar la imparcialidad en la libertad de la deliberación práctica depende precisamente de interpretar la libertad racional de un modo que se *aplicaría* también al pensamiento factual, como creía Kant.

[17] Al analizar este grupo de cuestiones, me he beneficiado particularmente del trabajo de Thomas Nagel y de nuestros debates.

en primera persona, e implica un *yo* que tiene que ser más íntimamente el yo de mis deseos de lo que esta concepción permite.

Cuando pienso en el mundo y me esfuerzo en encontrar la verdad con respecto a él, pienso *sobre el mundo,* emito juicios o planteo preguntas que hacen referencia al mundo y no a mí. Pregunto, por ejemplo,

¿Es el estroncio un metal?

o me digo sin vacilación

Wagner nunca conoció a Verdi.

Preguntas y afirmaciones de este tipo proyectan las sombras de la primera persona, tal como si

Me pregunto si el estroncio es un metal,

o

Creo que Wagner nunca conoció a Verdi.

Pero estas son contrapartidas derivativas, meramente reflexivas, de pensamientos que no me mencionan. Yo intervengo en ellos, por así decir, solo en el papel de quien tiene el pensamiento[18].

Por supuesto, también puedo aparecer en mis pensamientos de un modo más sustantivo e individual. Mis pensamientos pueden ser específicamente sobre mí, como en

¿Estoy enfermo?

Pensamientos de este tipo son sobre mí en un sentido en el que otros pensamientos que tengo no son sobre mí, sino sobre otra persona o cosa. Más interesante resulta reparar en el hecho de que yo puedo aparecer en mis pensamientos como el *locus* de la evidencia, como en

Me parece azul.

---

[18] Esto es parte de lo que sostiene Kant en los «Paralogismos» de la *Crítica de la razón pura*. Pero este *yo* trascendental, que es formal en el caso del pensamiento general, para Kant realiza una función mucho mayor en relación con la moral.

En tal caso, intervengo específicamente como yo mismo, y mis propiedades psicológicas efectivas son relevantes (así, dada la calidad de mi sentido de la vista, el hecho de que me parezca azul puede ser un indicador fiable de que es verde). Si pregunto

¿Qué pienso acerca de esta pregunta?

en un sentido también implica una referencia específica a mí mismo, con mis propiedades psicológicas efectivas; puede ser una invitación que me hago a descubrir mis creencias, como podría descubrir las de otra persona (aunque no sea exactamente del mismo modo). Pero

¿Qué he de pensar sobre esta pregunta?

en tanto que produce el mismo efecto que

¿Cuál es la verdad sobre esta pregunta?

es, de nuevo, un caso en el que *yo* intervengo solo derivativamente: la última pregunta es la primordial.

Debido a esto, un *yo* como este es también impersonal. La pregunta,

¿Qué he de pensar sobre esta pregunta?

podría ser también

¿Qué ha de pensar cualquiera sobre esta pregunta?

Esto es así incluso cuando significa

¿Qué he de pensar sobre esto en virtud de los datos de que dispongo?

Aquí se pregunta por lo que cualquiera habría de pensar sobre ello en virtud de esos datos. Igualmente, lo que cualquiera cree verdaderamente tiene que ser coherente con lo que los demás creen verdaderamente, y cualquiera que delibere sobre la verdad está comprometido, por la naturaleza del proceso, con el objetivo de tener un conjunto coherente de creencias, tanto con las propias como con las ajenas[19].

---

[19] Sin embargo, puede no ser posible unir cualquier verdad con cualquier otra, a causa del carácter perspectivista de cierto conocimiento; véase el capítulo 8.

No sucede lo mismo con la deliberación conducente a la acción. La deliberación práctica se lleva a cabo siempre en primera persona, y la primera persona no es derivativa ni puede ser reemplazada, de manera natural, por *nadie*. La acción que decido realizar será mía, y (en línea con lo que se dijo antes sobre los objetivos de la acción) el hecho de que sea mía no significa solo que se llegará a ella mediante esta deliberación, sino también que implicará cambios en el mundo de los cuales yo seré empíricamente la causa, y estos deseos y esta deliberación serán, por lo menos en parte, la causa. Es verdad que puedo distanciarme de mis deseos y reflexionar sobre ellos, y esta posibilidad puede considerarse, de hecho, como parte de la libertad racional a la que aspiran los agentes racionales. Esto excede en cierta medida las consideraciones anteriores sobre la libertad y la intencionalidad, pero no da aún el resultado requerido en relación con la moral. Al *yo* de la deliberación práctica reflexiva no se le exige que tenga en cuenta como dato el resultado de la deliberación llevada a cabo apropiadamente por otra persona, ni se le exige que se comprometa desde el principio con la armonía de las deliberaciones de todos; es decir, que establezca la regla desde un punto de vista igualitario. La deliberación reflexiva sobre la verdad introduce, de hecho, un punto de vista que es imparcial y que busca la armonía, pero esto es porque busca la verdad, no porque sea producto de una deliberación reflexiva, y estos rasgos no los compartirá la deliberación sobre qué hacer por la simple razón de que esta también sea reflexiva. El *yo* que en la reflexión racional se distancia de mis deseos es aún el *yo* que tiene esos deseos y que, empírica y concretamente, actuará; el mero hecho de distanciarse en la reflexión no lo convierte en un ser cuyo interés fundamental sea la armonía de todos los intereses. Este paso atrás no basta, por sí solo, para adquirir las motivaciones de la justicia.

Es realmente difícil explicar por qué el yo reflexivo, concebido como libre de compromiso alguno con todo deseo particular, habría de preocuparse por la satisfacción de cualquiera de estos deseos. El yo reflexivo de la deliberación teorética o factual tiene una unidad de interés con la creencia prerreflexiva: cada cual aspira a la verdad a su manera, y es por esto que la disposición prerreflexiva a creer genera fácilmente, en el caso normal, reflexiones correctivas. Pero en el modelo que estamos considerando aquí no hay ninguna identidad de intereses entre el yo práctico reflexivo y ninguno de los deseos particulares, míos o de los demás. No está claro, entonces, por qué el yo reflexivo habría de intentar satisfacer estos deseos. Estamos solo ante otro aspecto del error que

supone equiparar, como hace este argumento, la reflexión y el distanciamiento.

Siguen sin haberse respondido ciertas cuestiones profundas sobre lo que significa adoptar una perspectiva imparcial cuando uno *responde* a las motivaciones de la justicia. ¿Cómo un *yo* que adopta la perspectiva de la imparcialidad puede tener tanta identidad como para vivir una vida que respete sus propios intereses? Si la moral es posible, ¿me permite ser alguien en particular? Estas son cuestiones importantes en relación tanto con la moral como con la vida: respecto a la moral porque, como aspecto particular de lo ético, plantea la pregunta de un modo particularmente agudo, y respecto a la vida porque, desde cualquier concepción de las cuestiones éticas, se plantean desafíos reales referidos a las relaciones entre, por un lado, la imparcialidad y, por el otro, las satisfacciones y los objetivos individuales —o, de hecho, los compromisos personales, los cuales, sin ser egoístas, son más limitados que los impuestos por un interés o respeto universal por los derechos—. Algunas de estas cuestiones surgirán más adelante. Tienen que ver con lo que les ocurre al deseo y la deliberación personales cuando les influye la imparcialidad, en la medida en que alguien pueda alcanzarla. Espero haber mostrado en este capítulo que no hay ninguna ruta que nos lleve al punto de vista imparcial partiendo de la sola deliberación racional.

Capítulo 5

Estilos de teoría ética

El punto de vista imparcial puede ser invocado para un propósito diferente: no para persuadir a alguien de que realice el tránsito completo de la mera razón práctica a estar concernido por la justicia o la benevolencia, sino para apoyar o reclamar la prioridad de ciertas concepciones éticas sobre otras. En este caso, la cuestión es: atendiéndonos a personas que, en un sentido general, ya reflexionan teniendo en cuenta consideraciones éticas, ¿cómo han de pensar? ¿Son sensatas sus reflexiones éticas?

No me ocuparé aquí de la variedad de críticas dirigidas a las actitudes y las creencias éticas que se han esgrimido. Hay muchos estilos de crítica y los más convincentes se apoyan, como siempre ha ocurrido, no tanto en desplegar argumentos filosóficos como en mostrar que esas actitudes descansan en mitos, en falsedades acerca de la manera de ser de las personas. Incluso entre las críticas que entrañan más claramente la argumentación filosófica, no todas me interesan directamente. Algunos de estos modelos argumentativos tienen una utilidad puntual: sacar a la luz las consecuencias de las posiciones éticas o probar su incoherencia. Son herramientas para la argumentación ética. En este capítulo y en el siguiente me ocupo de un tipo de estructura más elaborada, completa y ambiciosa: la *teoría ética*. (Posteriormente consideraré la idea de que ciertas herramientas para la argumentación ética, si se aplican meticulosamente, son suficientes en sí mismas para generar una teoría ética).

¿Qué es una teoría ética? Una definición más bien compleja nos permitirá captar mejor el uso más provechoso de esta expresión: una teoría ética es una caracterización teórica del pensamiento y la práctica éticos que implica o bien un test general sobre la corrección de las creencias y los principios éticos básicos o, alternativamente, que no puede existir un test de este tipo. Aquí me ocuparé del primer tipo de teoría ética, el tipo positivo. En el presente capítulo, propondré una caracterización de dos estilos predominantes de teoría ética positiva, y en el siguiente trataré las motivaciones más profundas de las teorías de este tipo y su relación con la práctica. Previamente, sin embargo, tengo que decir algo acerca del tipo negativo de teoría ética y de por qué propongo este tipo peculiar de definición.

No importa mucho qué uso le demos a la expresión «teoría ética» siempre que lo explicitemos. Hay una razón, sin embargo, para usarla como propone esta definición, que comporta una cuestión filosófica significativa: hace veinte o treinta años era habitual distinguir las teorías «éticas» de las «metaéticas». Las primeras hacían afirmaciones sustantivas sobre qué deberíamos hacer, cómo habríamos de vivir, qué vale la pena, etc. Las segundas estudiaban el estatus de esas afirmaciones: si podían constituir conocimiento, cómo podían ser validadas, si eran objetivas (y en qué sentido), etc. La idea subyacente a esta distinción era, lógicamente, que los dos tipos podrían disociarse y que una teoría metaética, como tal, no tendría implicaciones éticas.

Es importante separar la propuesta de hacer esta distinción de otras dos ideas con las que a menudo se la vincula. Una es la de que el estudio metaético habría de ser lingüístico: una investigación de los términos que usamos en el discurso ético. Esto implica una precisión suplementaria sobre la naturaleza de la filosofía que no ha sido muy fructífera en filosofía moral. Aunque la distinción entre ética y metaética es independiente de la idea de que la metaética es de orden lingüístico, la formulación lingüística probablemente la favoreció, en virtud de la tesis de que tiene que ser posible distinguir entre los medios que un lenguaje nos proporciona para decir ciertas cosas y lo que decimos con él. Esta convicción estaba muy extendida en aquel momento y dio lugar a otras distinciones —como la distinción entre analítico y sintético— que hoy en día no gozan de gran predicamento. Pero aunque las preocupaciones de la filosofía lingüística favorecieran la distinción entre lo ético y lo metaético, esta distinción no se compromete con una formulación lingüística.

Otra idea que es independiente de la distinción, incluso cuando adopta una forma lingüística, es la de que la filosofía no habría de contener ninguna aseveración ética y limitarse a lo metaético. Esta estrategia descansa, sin duda, en otros presupuestos, relativos, una vez más, a la naturaleza de la filosofía, que no son universalmente compartidos. Así, Moore, quien en los *Principia Ethica* estableció una distinción tajante que resultó muy influyente entre decir qué es el bien y decir qué cosas son buenas, se permitió hacer en ese libro lo segundo tanto como lo primero, y aunque la distinción haya tenido una gran influencia entre ciertos filósofos, fue su teoría de qué cosas son intrínsecamente buenas la que más impresionó a otros filósofos. Moore tenía, no obstante, una concepción del bien y de nuestro conocimiento de él, que habría de permitir que la filosofía intentara decir qué cosas eran buenas. Pensaba que el bien se detectaba mediante un tipo de discriminación intelectual, y al menos parte de este procedimiento (que definió tan vagamente que no deja claro en qué medida) se parece tanto al análisis intelectual que hace inteligible que la filosofía, o las capacidades de los filósofos, pudiera contribuir en algo a ello. Pero quien piense que la tarea de la filosofía es en primer lugar el análisis, y que hacer juicios éticos sustantivos comporta cosas bastante diferentes del análisis intelectual, no hallará razón alguna por la que tales juicios hayan de ser parte de la filosofía y tratará de desterrarlos de ella, como algunos filósofos hacían hace veinte o treinta años.

La distinción entre lo ético y lo metaético no resulta ya tan convincente o significativa. Hay distintas razones para ello, pero la más relevante aquí es que hoy en día resulta obvio (de nuevo obvio) que lo que se piensa sobre el objeto del pensamiento ético y de su propósito tiene que afectar a qué clase de pruebas de aceptabilidad o coherencia le son apropiadas; y el uso de tales pruebas tiene que afectar a todo resultado ético sustantivo. Por otra parte, usar determinadas pruebas y modelos de argumentación puede conllevar una concepción y no otra de lo que es el pensamiento ético. Una teoría que combina concepciones sobre qué es el pensamiento ético y cómo habría de llevarse este a cabo y que produzca consecuencias sustantivas por hacerlo de ese modo, es una teoría ética positiva.

No obstante, determinadas concepciones del contenido y la naturaleza de la ética excluyen la existencia de ese tipo de pruebas. La más extrema de estas concepciones dice que mantener una posición ética consiste simplemente en elegir una entre todas y atenerse a ella. A mi

modo de ver, hay buenas razones para que una concepción como esta sea también considerada una teoría ética, aunque negativa; si bien cabe distinguirla de una teoría sobre la naturaleza del pensamiento ético que *deja abierta* la cuestión de si estas pruebas pueden existir. Alguien podría estar convencido y seguro sobre la caracterización que cabe dar a lo ético, y ser escéptico con respecto a la posibilidad de que existan esas pruebas; y hay opciones más complejas que esta, según las cuales puede haber pruebas en ciertos contextos culturales y no en otros. Este es el tipo de caracterización que defiendo en este libro, y es sensato no llamarlo teoría ética. Las teorías éticas son proyectos filosóficos e implican que la filosofía puede determinar, positiva o negativamente, cómo habríamos de pensar con respecto a cuestiones éticas; en los casos negativos, con el resultado de que realmente no podemos pensar mucho en ética. Es esta opción negativa la que los filósofos tenían normalmente presente cuando decían en el pasado que la filosofía no podría determinar cómo hemos de pensar sobre temas éticos.

Por el contrario, yo quiero afirmar que sí podemos pensar sobre cuestiones éticas, y que podemos hacerlo de muchas maneras, a menos que nuestras circunstancias históricas y culturales lo hayan hecho imposible; pero la filosofía no puede decir gran cosa con respecto a cómo hemos de hacerlo. El propósito de usar la expresión «teoría ética» de la manera que propongo es sacar a la luz la similitud entre las teorías positivas y las negativas en cuanto a sus presupuestos implícitos relativos al papel de la filosofía. En este punto del argumento puede parecer una idea demasiado sutil, pero espero que acabe no pareciéndolo. Mi objetivo es llegar a una perspectiva diferente de la de cualquiera de esas teorías; una perspectiva que incorpora cierto escepticismo sobre la ética filosófica, pero un escepticismo referido a la filosofía más que a la ética.

En cuanto a las teorías éticas positivas, existen varios tipos, así como varios modos de clasificarlas, que establecen diferentes tipos de tipos. Ninguna clasificación es particularmente iluminadora, pero es útil distinguir entre dos estilos básicos, el *contractualista* y el *utilitarista*. La idea fundamental del contractualismo ha sido formulada por T. M. Scanlon, en conexión con su caracterización de la mala conducta moral: «Un acto es malo si su realización, dadas las circunstancias, fuera desaprobada por cualquier sistema de reglas para la regulación general de la conducta que no sería razonable que alguien pudiera rechazar como

base de un acuerdo general informado y no forzado»[1] (Scanlon y otros autores con los que dialogaré suelen hablar en términos de *la moral;* yo haré lo mismo a veces). Esta caracterización de la mala conducta va unida a una teoría particular relativa al objeto del pensamiento moral, o de qué son los hechos morales en última instancia. Según esta teoría, el pensamiento moral se ocupa de qué acuerdos podrían alcanzar las personas en circunstancias favorables, en las que nadie ignorara lo relevante ni fuera coaccionado. La teoría comporta igualmente una caracterización de la motivación moral. El motivo moral básico es «un deseo de ser capaz de justificar las propias acciones ante los demás sobre fundamentos que no sería razonable rechazar»[2]. Hay un parentesco evidente entre este complejo de ideas y las concepciones kantianas examinadas en el capítulo anterior. Ahora bien, no se trata de mostrar que todo agente racional tenga que ser un ciudadano legislador de una república nocional, sino de qué reglas serían aceptables por quienes se supone que tienen interés en alcanzar un acuerdo.

El utilitarismo, por el contrario, toma los hechos relativos al bienestar individual como el tema básico del pensamiento ético. Hay muchos tipos de utilitarismo, que difieren en cómo cabe evaluar el bienestar y en otras cuestiones: si, por ejemplo, es el acto individual lo que habría de estar justificado en términos de maximización del bienestar o es más bien alguna regla, práctica o institución. (Esto es lo que distingue el utilitarismo *directo* del *indirecto).* Todas las variantes coinciden en la agregación del bienestar[3], esto es, en sumar de alguna manera el bienestar de todos los individuos implicados (esta fórmula, incluso la palabra «implicados», plantea muchas dificultades).

Examinaré estos estilos de teoría ética con más detalle. Pero estos meros esbozos introductorios ya dan cierta idea de cómo podrían llegar a resultados distintos. Una diferencia se halla en la manera más natural de definir la *circunscripción* de la moral que tiene cada teoría, esto es, de

---

[1] Scanlon, «Contractualism and Utilitarianism», en Sen y Williams, *Utilitarianism and Beyond,* pág. 110. Estoy en deuda con diversas ideas de este artículo.

[2] *Ibid.,* pág. 116.

[3] Esta puede ser concebida como la característica definitoria del utilitarismo; véase Sen y Williams, *Utilitarianism and Beyond,* págs. 2-4, y Amartya Sen, «Utilitarianism and Welfarism», *Journal of Philosophy,* 76 (1979). Para ciertas distinciones entre formas de utilitarismo, véase J. J. C. Smart y Bernard Williams, *Utilitarismo. Pro y contra.* James Griffin hace un repaso de la bibliografía reciente en «Modern Utilitarianism», *Revue internationale de philosophie,* 141 (1982).

quién se ocupa el sistema en primera instancia. Para el contractualismo la circunscripción natural abarca a aquellos ante quienes es concebible que tratemos de justificar nuestras acciones; en la interpretación más simple, los otros agentes morales. Esto puede ampliarse a una preocupación por los intereses de aquellos que son incapaces de dar o recibir justificaciones (niños pequeños, por ejemplo, o discapacitados psíquicos). En tales casos, nos resulta natural pensar, como ocurre en el ámbito jurídico, en fideicomisarios que actúan en su nombre. En una ampliación ulterior, podrían considerase los animales, pero estos están más alejados de la circunscripción primaria. Podríamos esperar que el contractualismo preste a los animales una atención diferente de la que otorga a las relaciones morales entre personas. La idea de un contrato, incluso en esta forma esquemática mínima, siempre sitúa en primer lugar las relaciones de igualdad entre los agentes que son tanto los sujetos como los objetos del pensamiento moral.

El utilitarismo adopta otra orientación. Una de las interpretaciones más naturales del bienestar con la que se vincula (y que históricamente es la primera) es la que lo define como placer y ausencia de dolor, y la circunscripción natural del utilitarismo la constituyen todas las criaturas capaces de sentir placer y dolor. Esta base ha sido refinada por las elaboraciones modernas, y es más probable que la circunscripción se defina hoy en día en términos de quienes tienen preferencias o voliciones y pueden sufrir por la frustración de las mismas. En la mayoría de las versiones, también los animales son parte de la circunscripción primaria: de hecho, es más natural incluir a ciertos animales que a determinados humanos (como las personas moribundas, por ejemplo). Esta concepción apela a una motivación moral: la benevolencia. E introduce, al mismo tiempo, una disparidad entre los agentes morales, por un lado, y los beneficiarios de la moral, por el otro. La segunda clase es, desde el principio, más amplia que la primera. Esta es una característica del utilitarismo que procede de su bienestarismo.

Otra característica importante deriva del hecho de que es un tipo de consecuencialismo; juzga las acciones en términos de sus consecuencias[4]. Para toda forma de consecuencialismo, el valor ético reside, en último término, en los estados de cosas. (En el caso del utilitarismo,

---

[4] Para las complejidades que exigen el refinamiento de esta fórmula, véase «Una crítica del utilitarismo», en Smart y Williams, *Utilitarismo. Pro y contra,* sec. 2.

que es un consecuencialismo del bienestar[5], este valor se encuentra en las diferencias en bienestar que producen los estados de cosas). El resultado es que, para el utilitarismo, la agencia es secundaria: nuestra relación ética básica con el mundo, en tanto que agentes, es la de ser la causa de estados de cosas deseables o indeseables. Nuestra preocupación ética básica es conseguir, en la medida de nuestras posibilidades, que haya más bienestar o utilidad en el mundo y no menos, y, en la versión más simple del utilitarismo, hemos de actuar simplemente de la manera que sea más eficaz para que esto ocurra. Es cuestión de qué palancas causales podamos accionar en ese momento. A veces, las conexiones causales por medio de las cuales puedo influir en los resultados dependen de las acciones de otras personas[6], pero esto no supone ninguna diferencia específica. Se trata sencillamente de saber qué cambios producen mayor bienestar. Lo que significa que hay estados de cosas que, debido a que puedo influir en ellos en lo que respecta al bienestar que producen, resultan ser de mi incumbencia cuando, sobre presupuestos no utilitaristas, serían de la incumbencia de otra persona. Además, debido a que la clase de los beneficiarios es mayor que la de los agentes, hay situaciones que son de la incumbencia de alguien aun cuando, desde una perspectiva no utilitarista, no lo serían de nadie.

Estas consideraciones sacan a relucir una diferencia adicional entre el utilitarismo y el contractualismo —referida, de nuevo, a las interpretaciones más inmediatas y naturales de ambas teorías—. Las exigencias del utilitarismo de maximizar la producción del bienestar son ilimitadas. No hay límite a lo que una persona dada podría hacer para mejorar el mundo, excepto los límites del tiempo y la fortaleza personal. Además, dado que las relaciones de los estados de cosas posibles con las acciones de cualquier persona son indeterminadas, las exigencias son ilimitadas en el sentido adicional de que no suele haber límites claros entre lo que se espera de mí y lo que se espera de los demás. Los teóricos utilitaristas proceden (con diferentes grados de entusiasmo)[7] a *restable-*

---

[5] En sentido estricto, se trata del consecuencialismo del bienestar que ordena los estados de cosas en función de su utilidad. Véase la nota 3 anterior.

[6] Para una discusión de algunos de los efectos que esto tiene en cuestiones relativas a la responsabilidad, véase «Una crítica del utilitarismo», secs. 3 y 5; Samuel Scheffler, *The Rejection of Consequentialism* (Nueva York, Oxford University Press, 1982).

[7] Algunos autores utilitaristas aspiran a hacer crecer un sentimiento de culpa indeterminado en sus lectores. Peter Singer es un ejemplo claro de ello; en su libro *Practical Ethics* (Nueva York, Cambridge University Press, 1980) [*Ética práctica*, trad. de R. Herrera

cer restricciones a lo que se le puede exigir a un individuo dado, diciendo que suele ser más eficiente que nos ocupemos específicamente de nuestros hijos, por ejemplo, o que descansemos ocasionalmente de realizar buenas obras. Los contractualistas, y muchos otros teóricos de cualquier adscripción, protestarán diciendo que esto invierte el orden de las cosas: no es potestad suya renunciar a mis derechos con respecto a mis hijos y a mi tiempo.

La idea de que existe un límite para «el rigor del compromiso», como dice John Rawls, es algo que contribuye a conformar su teoría, que es la concepción contractualista de la ética más rica y compleja que se haya propuesto hasta la fecha[8]. La teoría de la justicia de Rawls aspira a hallar principios que gobiernen la vida social y política más que la conducta individual. Pero parte de una base moral, y tiene también consecuencias importantes para el pensamiento puramente moral.

La teoría de Rawls es la elaboración de una idea simple: un sistema justo de convenciones es aquel en el que las partes pueden concordar sin saber cómo les beneficiará personalmente el acuerdo. La concreción de esta idea es la ficción conocida como la posición original, en la que las personas eligen principios sociales tras el «velo de la ignorancia», que les oculta sus posiciones sociales futuras, así como sus gustos e intereses individuales. No quedan ocultas las proposiciones generales, tales como los descubrimientos de las ciencias sociales, de modo que tienen cierta información con la que operar, pero no datos que les permitan hacer diferencias a su favor. La ignorancia es, pues, menos radical que la implicada en el capítulo anterior, en el que me ocupé de la idea kantiana de que las personas eligen como agentes racionales «y nada más». (Una consecuencia relevante del modelo es que Rawls está obli-

---

Bonet (Madrid, Akal, 2009)], está sin duda más interesado en producir este efecto que en proporcionar una base teórica para el mismo, cuestión que recibe un tratamiento muy superficial. Como persuasión moral, es probable que este tipo de táctica sea contraproducente y lleve a una preocupación moral defensiva y llena de resentimiento. Esto puede apreciarse en la bibliografía y, actualmente, en lo que ocurre a nuestro alrededor. Para el caso de la investigación, véase James S. Fishkin, *Beyond Subjective Morality* (New Haven, Yale University Press, 1984).

[8] Véase John Rawls, *Teoría de la justicia*, y los artículos citados en la nota 10 de este capítulo. Norman Daniels (ed.), *Reading Rawls* (Nueva York, Basic Books, 1975), es una compilación de artículos útil.

gado a asumir, de manera poco plausible, que el conocimiento de la historia no es esencial para la comprensión científica de la sociedad, a no ser que considere que alguien puede conocer el curso histórico sin saber cuál es su propio lugar en él, lo cual resulta aún menos plausible. Un aspecto importante de esta teoría, y de todas las que se le parecen, es su carácter radicalmente ahistórico). En tales circunstancias, las partes tienen que elegir la disposición social sobre la base de sus propios intereses. Su posición para hacerlo no es buena, dado que no saben quiénes son, pero eso no significa que su elección excluya principios benevolentes o altruistas. No hay que interpretar la teoría de Rawls como un intento de llegar a la justicia social a partir del interés propio. Su originalidad reside en presentar una elección interesada en condiciones de ignorancia de la propia identidad como modelo, en aspectos significativos, de una elección desinteresada o moral en condiciones normales de conocimiento.

El resultado de la deliberación en la posición original es que las partes aceptan dos principios fundamentales de la justicia:

(1) Cada persona ha de tener un derecho igual a la mayor libertad posible que sea compatible con una libertad semejante para los demás.
(2) Las desigualdades sociales y económicas han de ser estructuradas de manera que sean (a) para mayor beneficio de los menos aventajados, y (b) estén vinculadas a empleos y cargos accesibles a todos en condiciones de justa igualdad de oportunidades[9].

El segundo de estos principios se basa en la idea de que las partes harán uso en sus deliberaciones de una regla «maximin», esto es, una regla que jerarquiza las alternativas según sus peores resultados posibles. Este es un rasgo distintivo de la teoría de Rawls. Se supone que la elección de este principio en la posición original no descansa en ningún sesgo particularmente conservador de las partes (que no pueden permitirse ninguna preferencia en un sentido u otro), sino en el carácter peculiar de la elección, consistente en que a las partes no les es posible realizar una evaluación de probabilidades, no tienen gran interés en

[9] Rawls, *Teoría de la justicia,* págs. 60, 83; hay una presentación final, más elaborada, de los principios en la pág. 302 [1.ª ed. inglesa; págs. 82, 105 y 340-341, respectivamente, de la ed. castellana].

obtener un beneficio superior al mínimo, y los peores resultados implican «graves riesgos» que nadie podría asumir.

Esto expresa algunas ideas importantes sobre los bienes fundamentales; por ejemplo, que la esclavitud es simplemente inaceptable, con independencia de los beneficios que acarree. De hecho, como muestra su rechazo de todo cálculo de probabilidad, Rawls está comprometido con la conclusión de que ni siquiera una sociedad con un número reducido de esclavos sería aceptable. Esta puede ser una consecuencia positiva de una teoría moral de la justicia, pero no se sigue naturalmente del modelo de elección racional en condiciones de ignorancia. Si se trata de una elección racional basada en el interés propio, es difícil imaginar cómo podría evitarse la cuestión de las probabilidades, o por qué, si la probabilidad de acabar siendo esclavo fuera mínima, no sería racional que las partes escogieran un sistema que contempla la esclavitud si comporta una cantidad suficiente de beneficios de otro tipo. Por este tipo de razones, se ha criticado mucho el elemento de teoría de la decisión o de elección racional que incorpora el modelo de Rawls.

Se plantea también la importante cuestión de la naturaleza de los bienes en términos de los cuales se supone que las partes efectúan sus elecciones racionales. La teoría de la elección racional, en sus usos habituales, suele funcionar sobre la base de la utilidad o el bienestar individual, que, a su vez, están en función de las preferencias y los gustos de los agentes (veremos más adelante que, en manos de los utilitaristas, el bienestar no es un asunto tan sencillo). Pero las partes contratantes no tienen preferencias o gustos individuales conocidos de ningún tipo, de modo que esto no les es accesible. Rawls hace elegir a las partes sobre una lista de «bienes primarios», que son la libertad y las oportunidades, los ingresos y la riqueza y las bases del autorrespeto. Estos bienes vienen dados por lo que él llama «la teoría tenue del bien»; la idea es que son bienes que toda persona querrá si es que quiere algo. Pero la complejidad del asunto es mayor. No parece plausible que la lista de bienes primarios haya sido establecida solo porque se ha estimado que son los únicos necesarios para motivar la acción. Una consideración de este tipo no nos permite deducir más que la libertad. Es difícil concebir, también, cómo las partes podrían evitar la reflexión (que les es accesible por su conocimiento de hechos sociales generales) de que algunos de esos bienes primarios, en especial el dinero, son más importantes en algunas sociedades que en otras.

¿Por qué se introducen los bienes primarios? No es técnicamente imposible que las partes desprovistas de preferencias particulares elijan un tipo de sociedad sobre otro. Podríamos suponer que estas, sean quienes sean en el futuro, eligen una situación en la que, de aquello que prefieren, consiguen más (en esa situación) de lo que conseguirían en otras situaciones. Como veremos, esto es lo que hace R. M. Hare al construir su teoría utilitarista. Tales comparaciones pueden parecer fantasiosas, pero no es esta la objeción de Rawls a su uso en la posición original. Más bien, Rawls ha rechazado que se permita a las partes pensar solo en términos de lo que preferirían en las diversas situaciones sociales. Las partes son reacias, como lo somos nosotros, a considerar aceptable una situación solo porque la encontrarían aceptable si estuvieran en ella (hicimos notar antes nuestra propia renuencia con respecto a los «intereses auténticos»). Así pues, por mucho que las ciencias sociales nos dijeran, aunque es realmente improbable que lo hagan, que la mayoría de esclavos están contentos de librarse de la libertad, esto no daría a las partes razón alguna para elegir la esclavitud: desde el punto de vista de las personas libres que eligen libremente un plan de acción política, no es una opción. Rawls tenía razón, y estaba en su derecho al introducir estas convicciones en su teoría ética, pero la maquinaria de la elección racional en condiciones selectivas de ignorancia no las recoge de la mejor manera. Los bienes primarios pueden, quizá, entenderse mejor en términos de una concepción ética fundamental de la persona, y el mismo Rawls ha ido posteriormente en esa dirección[10].

Desde un punto de vista formal, el utilitarismo es una opción dentro del contractualismo. Si la pregunta contractualista no se plantea a propósito de principios o prácticas particulares, sino sobre un conjunto total de principios, como en el caso de Rawls, entonces las partes podrían escoger algún sistema utilitarista como respuesta. Dado el tipo de diferencias que ya hemos señalado en los resultados característicos de las dos formas de pensar, esto es improbable, pero el procedimiento no lo excluye. Consideremos que a ciertas personas se les pidiera que seleccionaran un conjunto de principios, armados de preferencias (en vez de un índice de bienes primarios), sin más restricciones en sus elecciones que estar sujetos a un grado de ignorancia que hace que no sepan

---

[10] Véase Rawls, «Kantian Construction in Moral Theory», *Journal of Philosophy*, 77 (1980); y «Social Unity and Primary Goods», en Sen y Williams, *Utilitarianism*. Véase también T. M. Scanlon, «Preference and Urgency», *Journal of Philosophy*, 72 (1975).

quiénes serán en el mundo gobernado por los principios que seleccionen. La idea de que determinados principios son *aceptables desde una perspectiva imparcial* se equipara a la de que estos son los principios que seleccionaría alguien convencido de tener las mismas opciones de ser tal o cual persona como resultado de su elección. (Puede verse ahora que basta una única persona para efectuar la elección). Este es el planteamiento de John Harsanyi, quien pretende suministrar un conjunto de principios que maximizarían la utilidad media de las personas afectadas. Esto se asemeja a un argumento contractualista, pero el resultado es utilitarista[11].

No obstante, en sus versiones más habituales, el utilitarismo parte de ideas relativas al bienestar o a los intereses de las personas. Su proyecto, que (en la forma más simple) consiste en considerar el bienestar de todos bajo el cálculo de diversos resultados alternativos, implica tanto problemas técnicos graves como dificultades conceptuales profundas. Una de ellas es que, excepto en los casos más simples, los resultados no afectan siempre a las mismas personas, y las personas afectadas por una alternativa podrían no existir en otras alternativas[12]. Pero no seguiré por la vía de las dificultades técnicas.

Dije anteriormente que para el utilitarismo el motivo moral característico es la benevolencia. Este es un término vago, que puede también llevar a equívoco, particularmente si sugiere sentimientos intensos de vinculación personal o bien cualquier tipo de sentimiento que se experimenta naturalmente con mayor intensidad hacia unas personas que hacia otras. La benevolencia utilitarista no implica vínculos particulares de ningún tipo y es inmune a la ley de la inversa del cuadrado. El término representa una relación positiva con los deseos y satisfacciones de otras personas que la persona benevolente siente solo porque se

---

[11] El argumento de Harsanyi aparece en algunos artículos incluidos en *Essays in Ethics, Social Behavior, and Scientific Explanation* (Boston, Dordrecht Reidel, 1976). Véase también «Morality and the Theory of Rational Behaviour», reimpreso en Sen y Williams, *Utilitarianism*. En «Contractualism and Utilitarianism», Scanlon hace una distinción interesante entre el argumento de Harsanyi y un planteamiento genuinamente contractualista.

[12] Esto tiene una especial relevancia para la política poblacional. Véase Derek Parfit, «On Doing the Best for our Children», en M. D. Bayles (ed.), *Ethics and Population* (Cambridge, Schenkman, 1976); «Future Generations: Further Problems», *Philosophy and Public Affaires,* 11 (1982); y *Razones y personas*. También vale la pena consultar J. McMahan, «Problems of Population Theory», *Ethics,* 92 (1981).

trata de los deseos de los demás y de su satisfacción. Esta idea general precisa de mayor concreción para poder desempeñar un papel en la teoría ética, y esta concreción la elaboran de manera diferente teóricos diversos. Al considerar la importante cuestión de la actitud imparcial del utilitarismo hacia la satisfacción de los deseos, examinaré una versión interesante del utilitarismo desarrollada por R. M. Hare, quien trata la relación del agente con los deseos ajenos (lo que llamé sin mucha precisión «benevolencia») en términos de identificación imaginativa.

La teoría de Hare[13] parte de ciertas tesis sobre la naturaleza de los juicios morales: son prescriptivos y son universales. «Prescriptivo» es un término relativo al lenguaje (del que me ocupé brevemente en el capítulo 4, y sobre el que tendré más que decir en el capítulo 7). Una emisión oral prescriptiva es del tipo «Hágase tal o cual cosa», y Hare considera que esa emisión, si es sincera, expresa un deseo o preferencia. Además, toda preferencia puede expresarse en forma de prescripción; por lo que cualquier agente que tenga preferencias está en posición de hacer prescripciones. Con todo, el agente no está aún comprometido a hacer prescripciones *universales;* las cuales se introducen con el lenguaje moral, en particular con los usos morales de *debe*. Hare no pretende, pues, que las premisas de todo razonamiento práctico impliquen una prescripción universal. Hasta aquí, Hare y yo estamos de acuerdo en que los compromisos del razonamiento moral pueden eludirse si el agente rehúsa el razonamiento moral.

El efecto de una prescripción universal, al juzgar que yo debo hacer tal cosa, es que acepto que cualquier otra persona debe actuar de modo semejante en circunstancias semejantes. En concreto, acepto que esto debería ser así si yo estuviese afectado por el resultado de la acción. Al considerar lo que debo hacer, por lo tanto, tengo que considerar lo que sentiría si yo fuera la persona afectada y, al hacerlo, aplico un «test de inversión de roles» y pienso qué es lo que yo querría o preferiría si estuviera en la posición del otro. Idealmente, habría de conducir este expe-

---

[13] Tal y como es presentada en *Moral Thinking* (Nueva York, Oxford University Press, 1981), libro en el que Hare ha desarrollado su teoría. Me referiré también a algunas de sus concepciones anteriores, expresadas en *The Language of Morals* (Nueva York, Oxford University Press, 1952; ed. rev. 1961) [*El lenguaje de la moral,* trad. de G. R. Carrió y E. Rabossi (México, UNAM, 1975)] y *Freedom and Reason* (Nueva York, Oxford University Press, 1965) [*Libertad y razón* (Barcelona, Crítica, 1997)].

rimento mental con respecto a toda persona, o incluso a toda criatura sentiente, implicada en situaciones semejantes.

El uso del test de inversión de roles no es exclusivo del utilitarismo. En una forma u otra, es un elemento fundamental del pensamiento ético, y una versión del mismo está implícita en el imperativo categórico kantiano. Los resultados distintivamente utilitaristas proceden de la interpretación específica y radical que da la teoría de Hare a la idea de situarnos en la posición del otro. Según esta interpretación, un agente se daría cuenta de lo que prefiere si se encontrase en esa posición solo si ahora adquiere una preferencia *real* correspondiente que se aplica a la situación hipotética. Me ocuparé más adelante de las razones de Hare para defender esta idea y de las críticas que se le pueden formular. En primer lugar, sin embargo, hemos de examinar el papel fundamental que desempeña en la transición al utilitarismo. Si idealmente alguien desarrollara un experimento mental completo cumpliendo con esas condiciones, adquiriría las preferencias correspondientes a las preferencias de cualquier otra persona implicada en la misma situación. Todas las preferencias se aglomerarían, así, en un único individuo. Pero, ¿cómo podría decidir el agente, dada esta aglomeración? Estas son ahora sus preferencias, y puede dotarlas de determinados requisitos racionales que supuestamente se aplican a cualquier deliberación en primera persona. Pero, sobre fundamentos éticos, no puede rechazar o descalificar ninguna preferencia que haya adquirido mediante la identificación con otras personas —ni siquiera las que él tenía inicialmente—, pues se supone que este nivel ideal de reflexión entraña una crítica de todo fundamento ético, y ninguno en concreto puede darse por sentado. Todo lo que tiene al agente idealmente reflexivo, a nivel ético, es el propio procedimiento de identificación aditiva. Así, una vez que las preferencias han sido ajustadas a la luz de criterios racionales que se aplican incluso a la deliberación en primera persona, no hay nada más que hacer, excepto comparar sus fortalezas relativas y elegir entre los diversos resultados sobre la base de esa comparación. El resultado es el utilitarismo.

Esta estructura es equivalente a una versión de lo que se conoce como teoría del observador ideal, que postula un observador omnisciente, imparcial y benevolente —que podríamos llamar el Agente Mundial—, que adquiere las preferencias de todo el mundo y las reúne. El test de lo que cabría hacer (o, en versiones indirectas de la teoría, de las prácticas o instituciones que cabría adoptar), se convierte, así, en el test de lo que elegiría ese observador. Como dice Hare, su modelo llega

al mismo resultado. Otra versión de la teoría del observador ideal excluye la condición de la benevolencia y no concibe un observador que asuma las preferencias de todos. Como señala Roderick Firth (quizá un tanto artificiosamente) en una exposición muy conocida de la teoría, el observador es «omnisciente, desinteresado, desapasionado, pero normal en todo lo demás». Esta versión de la teoría, en la que las diversas preferencias no se agregan en un Agente Mundial, sino que solo son detectadas externamente por un espíritu desapasionado, no conduce necesariamente al utilitarismo. El utilitarismo es solo uno de los candidatos que podría ser seleccionado por el test que propone la teoría. Sin embargo, en esta forma, la teoría del observador ideal afronta la objeción de que, si el observador no tiene ninguna motivación además de su imparcialidad, no hay razón para esperar que escoja algo; y, a menos que esa motivación sea benevolente —o que esté positivamente conectada con las preferencias que conoce—, podría tomar la decisión de frustrar tantas preferencias como sea posible[14].

En trabajos anteriores, Hare no hacía identificaciones hipotéticas tan exhaustivas, lo que le planteaba el problema de la posibilidad de que exista un «fanático» tan apegado a sus ideales que incluso aceptara los resultados hipotéticos desagradables del argumento de inversión de roles. Así, un nazi convencido podría aceptar la prescripción «Que se me mate si soy judío». Hare mantiene ahora que este problema se supera gracias a la misma línea de pensamiento que le llevó a la posición utilitarista. El nazi, si se implica en un proceso ideal de pensamiento, al identificarse con los judíos abandonará, en relación con las situaciones hipotéticas, sus preferencias antisemitas y adquirirá, para cada judío, una preferencia real contra el antisemitismo. Este proceso de pensamiento constituirá una crítica utilitarista (suponiendo que los cálculos se hacen bien) del antisemitismo. Se acepta, por supuesto, que un racista real puede rechazar adoptar este proceso de pensamiento, pero esto no socava el argumento.

---

[14] Hare tiene su propio observador ideal, llamado «el arcángel». Firth expone su propia teoría en «Ethical Absolutism and the Ideal Observer», *Philosophy and Phenomenological Research,* 12 (1952), págs. 317-345; y R. B. Brandt la critica en *A Theory of the Good and the Right* (Nueva York, Oxford University Press, 1979), págs. 225 y ss. Para una crítica de las dos versiones, véase Derek Parfit, «Later Selves and Moral Principles», en Alan Montefiore (ed.), *Philosophy and Personal Relations* (Montreal, McGill-Queens University Press, 1973), págs. 149-150 y notas 30-34.

Es sorprendente la fuerza que Hare atribuye al poder de la argumentación racional en ética. ¿Por qué se consideraba al fanático un problema primordial? Hare dice que «se demostraría que existe una laguna en las defensas del utilitarismo si [un fanático] pudiera escuchar, y comprender, todos los argumentos, y admitir todos los hechos, aducidos por un utilitarista, sin renunciar a su opinión»[15]. A primera vista, esto parece implicar una concepción desmesurada de la defensa, según la cual solo tenemos una defensa adecuada si somos capaces de aniquilar al adversario. Contra esto, sería sin duda posible mantener un conjunto racional, o razonable, de creencias éticas, y que otras personas sostuviesen creencias éticas diferentes que pudiéramos incluso deplorar pero no demostrar que fuesen incoherentes u objetivamente erróneas. (El mismo Hare, en su primera obra, *El lenguaje de la moral,* pensaba esto). Resulta más interesante el hecho de que podamos pensar que algunas posiciones que difieren de la nuestra son irracionales, y el racismo es una de ellas, pero que su irracionalidad no puede necesariamente demostrarse mediante los mismos argumentos en todos los casos, ni tampoco por medio de algún modelo de argumentación fundamental en la construcción de nuestras creencias éticas. Puede haber algo *específicamente* irracional en el racismo.

La equiparación que hace Hare entre la defensa y el ataque bebe de dos fuentes: uno es específico de su perspectiva, mientras que el otro es compartido con muchas otras teorías éticas. La idea general es que las teorías éticas de este estilo pueden concebirse fácilmente como armas ofensivas dirigidas contra el prejuicio, de modo que, si hay algún tipo importante de prejuicio que se les resiste, esto significa que no están bien diseñadas para realizar su función y es probable que hayan de ser reemplazadas, si no por el prejuicio, sí por otra teoría ética con más potencia de fuego. En cierta medida, esto es cierto de todas ellas, aunque difieran en la agresividad de sus ambiciones.

Hare basa esta concepción de las defensas en la creencia de que su argumentación a favor del utilitarismo se sigue estrictamente del significado de los términos morales. Habrá una laguna en las defensas *de esta tesis* si puede haber alguien que utilice de manera correcta el lenguaje moral y a la vez rechace la teoría; en tal caso, la tesis no solo quedaría sin defensa, sino que resultaría también refutada. Pero esta no es una

---

[15] Hare, *Moral Thinking,* pág. 170.

tesis razonable. No hay ninguna manera razonable de mostrar que las teorías alternativas malinterpretan o hacen un uso improcedente del lenguaje moral. Incluso si hubiera una característica básica del «lenguaje moral» en cuanto tal, e incluso si esta consistiera en que es prescriptivo y universal, no por ello nos veríamos impelidos ineluctablemente a suscribir la teoría. Hay otras interpretaciones de lo que supone aceptar una prescripción, y de lo que significa que algo sea universalizable, que nos conducirían a teorías diferentes. Como afirma John Mackie, hay diversos grados de universalización menos extremos que la última fase que representa la teoría de Hare, la cual iguala la totalidad de los gustos e ideales del agente en el experimento mental de la identificación. En palabras de Mackie, «como mucho, solo la primera fase, el apartamiento de las diferencias puramente numéricas como algo moralmente irrelevante, ha sido incorporada al significado del lenguaje moral»[16].

Los usos habituales del argumento de la inversión de roles, menos ambiciosos que el intento de hacer que fundamente todas las consideraciones morales, suelen incorporar los gustos personales en la identificación imaginaria, pero dejan fuera los ideales o las creencias éticas. Hacer esta distinción implica resistirse a una estrategia característica del pensamiento utilitarista que se conoce como *reducción* y se define como «el hecho de reducir todos los intereses, ideas, aspiraciones y deseos a un mismo nivel, y considerar que todos ellos pueden ser representados como preferencias, con diferentes grados de intensidad quizá, pero tratados igual en lo restante»[17]. Los teóricos utilitaristas a menudo comienzan pidiendo igual consideración para los intereses de todos, para luego extender esta petición hacia arriba (por así decir) hasta llegar a los ideales, y hacia abajo hasta incluir los meros gustos. Esta asimilación desfigura las proporciones de las cosas en muchos aspectos. En cierto sentido, subestima el significado de los ideales o de las concepcio-

---

[16] John Mackie, *Ethics: Inventing Right and Wrong* (Nueva York, Penguin, 1977), pág. 97 [*Ética: la invención de lo bueno y lo malo,* trad. de T. Fernández Aúz (Barcelona, Gedisa, 2000), pág. 111]. El «apartamiento —o la exclusión— de las diferencias puramente numéricas» se refiere al hecho de que el que yo sea yo y tú seas tú no puede contar en sí mismo como razón para tratar a uno de nosotros de modo diferente al otro. Dudo de que ni siquiera esto tenga nada que ver, específicamente, con el «lenguaje moral». Para profundizar en el tema de la universalización, véase el capítulo 6.

[17] Sen y Williams, *Utilitarianism,* pág. 8.

nes éticas, y exige al agente que abandone toda posición de principio o toda convicción profunda si un vasto agregado de preferencias, sean las que sean, favorece la acción contraria. La asimilación no otorga suficiente importancia al papel que tienen nuestras convicciones en los cálculos que realizamos. Simultáneamente, puede dar demasiada importancia a las convicciones de otras personas. Así pues, aunque los experimentos mentales de Hare nos proporcionan un argumento contra el fanático racista, este argumento no es del tipo adecuado.

En primer lugar, las sumas tienen que ser justas. Es característico del utilitarismo que sus resultados dependan de los cálculos, pero, en el contexto del racismo, este es un rasgo particularmente indeseable. Si el prejuicio racista es dirigido contra una pequeña minoría, por una mayoría que obtiene suficiente satisfacción de él, estamos a un tris de justificar el racismo. La cuestión no es la probabilidad de que este surja, o en qué circunstancias, sino que la pregunta en sí de cuántos racistas están involucrados no puede tomarse como una consideración aceptable con respecto a si el racismo es admisible (contrástese con el tratamiento que da Rawls a la esclavitud). Además, según el argumento utilitarista, se plantea como *una* consideración —aunque, si los cálculos son justos, no constituye una consideración decisiva o determinante— que los racistas obtienen cierta satisfacción del sufrimiento de los judíos. Pero esta no puede ser una consideración *en absoluto.* Ello no significa que los sufrimientos de los racistas no cuenten nunca. Significa que los sufrimientos que estos experimentan solo por mor de sus opiniones racistas no cuentan. Harsanyi añade en su sistema una cláusula para hacer frente a este tipo de problema: excluye del cálculo, de manera más bien brusca, las preferencias «antisociales». Pero no explica cómo definirlas. Las razones que da a favor de su cláusula sugieren que lo que está en cuestión son las preferencias antiutilitaristas[18].

Ya he mencionado los requisitos que establece la teoría de Hare, según los cuales las preferencias, cuando se reúnen en un agente, habrían de modificarse por referencia a criterios relativos a la racionalidad de primera persona. Lo que se toma en consideración, en último término, no son las preferencias efectivas de los agentes (incluyendo sus preferencias hipotéticas reales, esto es, las preferencias que tendrían si estu-

[18] En «Morality and the Theory of Rational Behaviour», citado en la nota 11.

vieran en las situaciones hipotéticas), sino más bien sus «preferencias totalmente prudentes», esto es, lo que preferirían si tuvieran una información completa y no cayeran en confusiones al razonar. Harsanyi propone una cláusula parecida. El pensamiento moral ha sido asimilado (de manera diferente en una y otra teorías) al pensamiento prudencial de un solo individuo. La conducta prudencialmente racional para un agente normal no se corresponde necesariamente con lo que él prefiere realmente, dado que puede estar confundido o mal informado. Ocurre lo mismo en el caso del conocimiento que tenemos de nuestras preferencias futuras. Habríamos de incluir entre nuestras preferencias «de ahora para después» (como las llama Hare, muy ilustrativamente) también la anticipación de nuestras preferencias «de después para después». La manera en que hacemos esto, según Hare, es tomando como sustitutos nuestras preferencias reales «de ahora para después», exactamente como hace el agente reflexivo con las preferencias de los demás[19]. Esto tiene muchas complicaciones. Solo me referiré al tipo de tratamiento que da el modelo de Hare a esas preferencias agregadas.

Este proceso de corrección de preferencias —su *idealización*— es apropiado para los modelos en los que todas las preferencias son nocionalmente las preferencias de una persona, como en el caso de la interpretación del Agente Mundial en la teoría del observador ideal. Sin embargo, solo es apropiado si el modelo se toma literalmente; pero, si se concibe literalmente, incluso en un grado mínimo, se hace patente su carácter estrambótico. Cualquier agente individual que tenga proyectos tan conflictivos, tan competitivos y con bases tan diversas como los del Agente Mundial se encontraría, cuanto menos, en una situación delicada. Le haría falta un conjunto de valores o deseos de segundo orden para poder ponderar la diversidad de sus preferencias. Pero, si el Agente Mundial los tiene, y son aún reconocibles en el agregado de preferencias, vuelve a tener demasiados. La verdad es que este agregado de preferencias es sencillamente ininteligible a menos que estas se conciban como preferencias de *personas diferentes*. El recurso al Agente

---

[19] Hare señala que este método en sí mismo no tiene por qué maximizar siempre el bienestar (*Moral Thinking*, págs. 101 y ss.). Que lo haga o no depende de otra cuestión, de cómo ponderemos las preferencias «de ahora para después», que sustituyen a preferencias «de después para después», con otras «de ahora para después» y «de ahora para ahora». Para un análisis sutil del interés prudencial por el futuro y sus analogías con el interés por otras personas, véase Parfit, *Razones y personas*.

Mundial nos obliga a olvidar ese hecho, nos obliga a ver el mundo ético como un océano de preferencias. Es lo que hacen, en diversos grados, todas las formas de utilitarismo. Esta idea suele criticarse por sus resultados éticos, pero la objeción fundamental es que no tiene sentido como interpretación del mundo. Es por ello por lo que no proporciona comprensión ética. «Que las personas son entes separados», como afirma John Findlay, constituye «el hecho fundamental de la moral»[20].

Si el utilitarismo se interpreta en términos menos drásticos que los del modelo del Agente Mundial, la idealización de las preferencias resulta menos apropiada. Ciertamente, no es nunca apropiada *meramente* porque las preferencias de un agente estén basadas en una información falsa. Será apropiada si, como resultado de la acción a que dan origen esas preferencias, el agente en cuestión y los demás obtienen menos utilidad de la que obtendrían de la acción basada en una corrección de preferencias. Para tomar un ejemplo político, si una administración utilitarista opera no sobre la base de lo que prefieren las personas, sino sobre la base de lo que preferirían si estuvieran mejor informadas, es posible que los administrados estén siempre descontentos con lo que hace la administración, puesto que no se libran nunca de sus errores y, por consiguiente, nunca tendrán realmente la preferencia idealizada para cuya satisfacción está diseñada la acción política.

El dudoso papel de la idealización en la teoría utilitarista está conectado con el procedimiento de reducción que he mencionado antes, en el que los intereses se asimilan a las preferencias. La idealización o corrección es apropiada cuando uno piensa en los intereses de las personas; uno de los hechos básicos a este respecto es que las personas pueden equivocarse con respecto a sus intereses. Pero la cuestión no es si podemos corregir apropiadamente las preferencias de los demás al reflexionar sobre sus intereses, sino en qué medida tenemos el derecho a actuar sobre la base de esas correcciones si las personas implicadas no las reconocen. Por el contrario, si nos preocupamos solo de saber qué grado de satisfacción de preferencias contiene el mundo, la cuestión será otra: si la idealización en un caso dado dará lugar a más utilidad a largo plazo. Se puede ver que es natural que estas dos cuestiones diferentes figuren en dos concepciones distintas de la política.

---

[20] John Findlay, *Values and Intentions* (Atlantic Highlands, N. J., Humanities Press, 1978), págs. 235-236.

La idealización está relacionada de forma ambigua con el test de la inversión de roles. Como hemos visto, el test puede ser apropiado si se aplica a las deliberaciones del Agente Mundial, después de que este haya hecho suyas todas las preferencias. Aun así, este resultado entra en conflicto con el espíritu con el que se nos recomendó inicialmente el experimento mental de la identificación con los demás. La pregunta originaria era «¿Cómo sería para mí estar en su lugar?». Esto lo interpreta Hare como equivalente a «¿Cómo sería para él?»; en la transferencia se pierde completamente el *mí* original. (Lo hipotético se entiende igual que en la conocida respuesta al comentario de que «Si yo hubiera sido Roosevelt, no habría hecho esas concesiones a Stalin»: «No seas estúpido, si tú hubieras sido Roosevelt, habrías hecho todo lo que Roosevelt hizo»). Pero, si el grado de identificación tiene que ser total, se sigue que, si las preferencias de otra persona están equivocadas, las preferencias que imagino que yo tendría también están equivocadas, y si de lo que se trata es de la *identificación,* entonces las preferencias han de continuar estando equivocadas. El resultado es, sin embargo, que identificándome totalmente con los sentimientos de otra persona acabo mejorando sus preferencias. Esta es una ilustración sumaria de una verdad sobre toda política utilitarista: la benevolencia extrae sus credenciales de la simpatía y las transmite al paternalismo.

Llegados a este punto, deberíamos recordar las razones que daba Hare a favor de la interpretación que convierte a cada persona idealmente reflexiva en una versión del Agente Mundial. Sus raíces se encuentran en las conexiones que Hare establece entre dos proposiciones que, manifiestamente, son muy diferentes: (1) Ahora prefiero con la intensidad S que, si yo estuviera en esa situación, sucediera X más que todo lo contrario. (2) Si yo estuviera en esa situación, preferiría con la intensidad S que sucediera X más que todo lo contrario. «Lo que afirmo», dice Hare, «no es que esas proposiciones sean idénticas, sino que no puedo saber que (2), ni lo que ello supondría, sin que (1) sea verdadera»[21]. La tesis de Hare, en definitiva, se refiere al conocimiento: no puedo saber ahora que en determinada situación preferiría con determinada intensidad que sucediera X, a menos que ahora prefiera con la misma intensidad que suceda X en esa situación.

---

[21] Hare, *Moral Thinking,* págs. 95-96.

Esta tesis parece difícil de aceptar incluso cuando el *yo* de la situación hipotética es claramente el *mío*, como ocurre en la toma de un seguro o en otras decisiones prudenciales semejantes. En este momento sé, por ejemplo, que si se quemara mi casa preferiría, con la mayor intensidad posible, que mi familia y yo pudiésemos escapar de las llamas. Dado que soy un agente moderadamente racional, realizo ciertas acciones ahora para asegurarme de que podríamos hacerlo si la situación se presentara, y esa acción procede sin duda de una preferencia que tengo en este momento. Pero esta preferencia prudencial presente no tiene de ninguna manera la misma intensidad que la preferencia que tendría si la casa estuviera realmente en llamas (lo que eliminaría casi cualquier otra consideración de mi mente), y no sería racional que la tuviera. No en balde, la intensidad de la preferencia derivará en parte de la probabilidad de la situación imaginada.

En el caso del experimento mental que acompaña a la versión radical del test de la inversión de roles, podría aducirse que la probabilidad de la situación imaginada es siempre cero, dado que se trata de la probabilidad de ser otra persona. Pero esta réplica es injusta, pues el experimento tiene que ver con situaciones descritas en términos generales, por lo que es cuestión de satisfacer cierta descripción general y no de ser otro individuo. Sin embargo, la probabilidad en muchos casos será también cero; y de todas formas no se presupone la intervención de la probabilidad en el argumento. Si concedemos esto, parece incluso menos plausible que la preferencia derivada pueda tener la misma intensidad que la preferencia que existiría en el caso de que la situación se produjese realmente. De hecho, no está ni tan siquiera claro qué significa decir que tiene o no determinada intensidad, dado que, en general, no hay un test independiente de la intensidad de la preferencia derivada.

Esto no quiere decir que no haya preferencias, con grados diversos de intensidad, basadas en la identificación empática con los demás. Por supuesto, las hay y son fundamentales para la experiencia ética. La cuestión es que la comprensión, la identificación y la preferencia no están relacionadas entre sí del modo que sugiere el modelo del Agente Mundial. Ante alguien que se encuentra en una situación de emergencia extrema, si soy una persona con humanidad, me formaré una preferencia predominante de socorrerla si puedo. Esta preferencia opera por medio de la consideración del estado en el que se encuentra alguien; una consideración en la que desempeña un papel esencial el pensa-

miento de lo que sentiría yo en esa situación o en una situación análoga. Lo que sé de lo que desea otro (que le ayude a escapar del fuego, por ejemplo) suscita en mí, debido a una disposición humanitaria, el deseo de ayudarle a escapar del incendio. Esta situación comporta, pues, cuatro verdades que me conciernen. En primer lugar, sé lo que él siente y que desea ser socorrido. En segundo lugar, sé que si yo estuviera en esa situación desearía recibir ayuda. En tercer lugar, ahora tengo la preferencia, en mi propia persona, de ser socorrido en una situación semejante. En cuarto lugar, al tener una disposición humanitaria, quiero socorrerla. En el modelo de Hare, la primera verdad es equivalente a la segunda, porque se entiende que *estar en esa situación* significa la inmersión total. (Esto revela hasta qué punto es radical esta interpretación. En la vida cotidiana no se consideran equivalentes estas dos verdades, y muchas situaciones cómicas son posibles gracias a su confusión). Para el modelo de Hare, además, no puedo conocer la segunda a no ser que la tercera sea cierta; y, por último, «tener una disposición humanitaria» significa estar dispuesto a hacer cálculos racionales en primera persona en los que las preferencias derivadas de ese tipo adquieren una importancia adecuada (contra mi propia conveniencia, etc.).

Todo esto no puede ser correcto. A lo largo de la historia de la filosofía moral se ha debatido mucho sobre cómo opera la comprensión por simpatía —o, como suele denominarse ahora, la «empatía»— y se han elaborado interpretaciones muy diversas. Pero una cosa que tiene que ser cierta es que la comprensión perspicaz de los sentimientos ajenos que posee la persona dotada de simpatía también la posee, en una forma muy parecida, el sádico o la persona cruel; esta es una de las maneras en que las personas crueles se distinguen de las brutales o de las indiferentes. Pero la persona cruel es alguien que no tiene, en absoluto, la preferencia de prestar ayuda (no es alguien que tenga una preferencia de ayudar, pero la vea superada por la preferencia de disfrutar del sufrimiento). Sin embargo, no hay duda de que *sabe*. Hare dice, sobre la vinculación que establece, que «es una verdad conceptual, en el sentido de "saber" que exige el pensamiento moral»[22]. Pero el pensamiento moral no exige ningún sentido de «saber» excepto el del conocimiento, y es una verdad, conceptual o no, que cualquier conocimiento del que haga uso el pensamiento moral puede volverse en su contra.

---

[22] *Ibid.*, pág. 96.

Me he detenido en esta cuestión porque el funcionamiento de la simpatía y del test de inversión de roles (que no son necesariamente la misma cosa) son importantes para el pensamiento ético; y también, de manera más inmediata, porque están implicados en la interpretación predominante del utilitarismo basada en el Agente Mundial. Este no es el único modelo al que puede recurrir el utilitarismo, y hay otras versiones que escaparán a algunas de estas críticas específicas. Pero la idea de asumir los deseos y sufrimientos del mundo y, a un nivel ideal siquiera, sentir todos los dolores y placeres como igualmente próximos a uno mismo, es una motivación básica del utilitarismo, que acentúa su contraste con el contractualismo; y la versión de Hare de lo que significa realmente el modelo del Agente Mundial hace patente con excepcional claridad lo que esto supone.

El utilitarismo es la más ambiciosa de las teorías éticas existentes. Aspira a alcanzar los resultados más definitivos y está dispuesto a imponerlos de la manera más firme posible contra las creencias éticas cotidianas. Tenemos que examinar, a continuación, la relación del utilitarismo y de otras teorías éticas con la práctica. ¿Por qué esas teorías habrían de estar investidas de alguna autoridad?

Capítulo 6

Teoría y prejuicio

Las teorías éticas tienen que adoptar un punto de partida u otro. Hasta el momento, he considerado cómo se podría partir desde fuera de la ética. También apunté la idea de comenzar desde dentro de la ética, pero meramente a partir del significado de los términos morales. Todas esas opciones me han parecido muy poco convincentes, y algunas las rechazo por completo. Muchos autores estarían de acuerdo con estas conclusiones, incluidos quienes aspiran a construir una teoría ética. Estos aún tienen que partir de algún lugar, y el único punto de partida que queda es la experiencia ética misma.

La «experiencia ética» puede abarcar muchas cosas. Una opción podría ser que la filosofía moral partiera de nuestra experiencia de la vida ética. Esta filosofía reflexionaría sobre lo que creemos, sentimos, presuponemos; los modos en los que confrontamos las obligaciones y reconocemos la responsabilidad; los sentimientos de culpa y de vergüenza, lo cual conllevaría toda una fenomenología de la vida ética. Esta podría ser una buena filosofía, pero sería improbable que desembocara en una teoría ética. Dado su interés por los tests, las teorías éticas tienden a partir de un único aspecto de la experiencia ética: las creencias. La manera más natural de concebir una teoría ética es considerarla como una estructura de proposiciones, que, al igual que una teoría científica, proporciona por un lado un marco para nuestras creencias y, por otro, las critica o las revisa. Así, su

punto de partida son nuestras creencias, aunque estas puedan ser reemplazadas.

Esas creencias éticas iniciales suelen denominarse *intuiciones* en la filosofía actual, pero este término ya no tiene las implicaciones de antaño. Por intuición solía entenderse una capacidad intelectual que nos permitiría llegar a verdades abstractas, y su aplicación a la ética respondía a la idea de que las verdades éticas podían captarse *a priori* por medio de esa facultad. Los filósofos que usaban este modelo[1] divergían en varias cuestiones: qué conceptos aparecían en las verdades dadas por la intuición (si era la bondad o la obligación, por ejemplo) o si esas verdades eran muy particulares o muy generales. Pero, al usar la noción de intuición, todos presuponían que nuestro modo de aprehender esas verdades éticas era significativamente semejante a como aprehendemos las verdades matemáticas, así como otras verdades necesarias. Las verdades éticas aprehendidas mediante la intuición podrían proporcionar un punto de partida para la teoría ética, en caso de haberla, pero no todos los partidarios de la intuición querían establecer una teoría ética, pues se suponía que la intuición misma proporciona el test o, más bien, hace innecesarios los tests.

Una serie de críticos han demolido este modelo de la intuición en ética[2], y las ruinas que se mantienen no son tan impresionantes como para que se escriba la historia de lo sucedido. Abreviando, lo que se le reprochaba era no poder explicar cómo una verdad eterna daba lugar a consideraciones prácticas, y que se equivocaba al asimilar las verdades éticas a las necesidades. Si verdades necesarias como las de las matemáticas fueran negadas por informantes de otras culturas buscaríamos, por supuesto, un traductor mejor, pero no sucede lo mismo con las creen-

---

[1] En el siglo xx, G. E. Moore, David Ross (véase el capítulo 10) y H. A. Prichard fueron intuicionistas. Entre los representantes anteriores de una tradición semejante, cabe destacar a Richard Price *(A Review of the Principal Questions in Morals,* 1788) y William Whewell *(The Elements of Morality,* 1845).

[2] John Stuart Mill atacó constantemente el intuicionismo, en particular por su dogmatismo; véase su *Autobiography,* ed. Jack Stillinger (Boston, Houghton Mifflin Co., 1964), págs. 134-135 [*Autobiografía,* trad. de C. Mellizo (Madrid, Alianza, 2008)]. La epistemología intuicionista fue duramente criticada en los años cincuenta. Aparte de *El lenguaje de la moral* de Hare, véase Stephen Toulmin, *The Place of Reason in Ethics* (Nueva York, Cambridge University Press, 1950) [*El puesto de la razón en la ética,* trad. de I. F. Ariza (Madrid, Alianza, 1979)] y Patrick Nowell-Smith, *Ethics* (Nueva York, Penguin, 1954) [*Ética,* trad. de G. Gutiérrez López (Madrid, Verbo divino, 1977)].

cias éticas. Sobre todo, se aducía que el recurso a la intuición como facultad no explicaba nada; se limitaba a decir que esas verdades eran conocidas, sin saber cómo. La «intuición» no es muy explicativa cuando se aplica a las verdades necesarias, pero con las creencias éticas es aún peor, por razones concernientes, una vez más, al desacuerdo cultural. Por muy poco que sepamos al respecto, ya sabemos demasiado sobre la explicación de las creencias éticas y de sus diferencias culturales como para aceptar un modelo que rechaza la posibilidad de que exista una explicación tal.

Así pues, en ética, la intuición, como facultad, se ha desvanecido. Pero las *intuiciones* —las creencias que presuntamente nos daba aquella facultad cuando se suponía que existía— son parte destacada de la disciplina. Se trata de convicciones espontáneas, moderadamente reflexivas sin ser fruto de la teorización, sobre la respuesta a cuestiones éticas, normalmente hipotéticas y planteadas en términos generales. Son a menudo preguntas sobre qué hacer. «¿Qué habríamos de hacer si, cambiando las agujas de una vía, pudiéramos desviar un tranvía de una vía, en la que con toda certeza acabaría con la vida de tres viejos, a otra vía en la que con toda certeza acabaría con la vida de un niño y un violinista de talento?». Este ejemplo no es mucho más extravagante que algunos de los que se han propuesto[3]. Pero las intuiciones no tienen por qué expresarse como respuestas a preguntas sobre qué hacer. Algunas pueden observarse en nuestra disposición a aplicar a cierta situación imaginada uno de esos conceptos éticos más sustantivos, tales como los que individúan virtudes o tipos de acción, que se mencionaron en el capítulo 1.

---

[3] Para ser justos, el dilema del tranvía, como se le ha dado en llamar, lo introdujo inicialmente Philippa Foot en «El problema del aborto y la doctrina del doble efecto» (reimpreso en *Las virtudes y los vicios)* para explorar la cuestión de hasta qué punto la estructura causal de una situación podría ser relevante para extraer conclusiones morales sobre esa situación; una cuestión que implicaba doctrinas tales como el principio del doble efecto. Algunos ejemplos pueden ser relevantes para ese tipo de cuestión aunque sean fantásticos. Puede darse el caso de que los ejemplos sean relevantes precisamente porque son fantásticos. Un caso destacado es el que usa Judith Jarvis Thomson en la discusión sobre el aborto, que es efectivo justo por su espantosa irrealidad; «In Defence of Abortion», reimpreso en Marshal Cohen, Thomas Nagel y Thomas Scanlon (eds.), *The Rights and Wrongs of Abortion* (Princeton, Princeton University Press, 1974). Para una crítica general del uso de ejemplos fantásticos en relación con las intuiciones morales cotidianas (aunque no en lo que se refiere al fundamento del pensamiento moral), véase Hare, *Moral Thinking*.

Existe una analogía que ha favorecido la recuperación del término «intuición» en estos contextos. Es con respecto al uso que se hace de él en lingüística y en filosofía del lenguaje para referirse a la comprensión espontánea del hablante de lo que puede y no puede decirse en su lengua, o de lo que puede decirse correctamente en una situación particular. Así, un angloparlante competente tiene la intuición de que no es correcto decir en inglés (como oí una vez afirmar a un filósofo del lenguaje de origen extranjero): «*In English we are not using the present continuous to signify a custom or practice*»*. Tales intuiciones son la materia prima de una teoría del lenguaje natural. Tenemos buenas razones para creer que sería posible formar esa teoría, dando cuenta de las reglas interiorizadas por el hablante, solo porque el hablante puede reconocer sin dudar como correctas o incorrectas en su lengua frases que nunca antes ha escuchado. Como ha señalado Noam Chomsky, esto es algo que hacemos constantemente. Además, algunos teóricos, especialmente Chomsky, creen que, dado que cualquier ser humano puede aprender de niño cualquier lengua humana, tenemos buenas razones para esperar que exista una teoría de las reglas que subyacen a todos los lenguajes naturales, una gramática universal.

¿Cómo se aplica esta concepción lingüística de la intuición a la ética? Hay un tipo de intuición pertinente para la ética que ciertamente se adecua al modelo, dado que es una mera aplicación suya. En el caso de los términos sustantivos referentes a las virtudes y a los tipos de acción, son posibles las intuiciones lingüísticas sobre las situaciones a las que cabe aplicarlos, justo porque son términos generales del lenguaje con condiciones complejas de aplicación. (Qué consecuencias éticas, si las hubiera, se siguen de la capacidad de las personas para usar tales términos, si bien difiere de una cultura a otra, es algo de lo que nos ocuparemos en los capítulos 8 y 9). En el caso de términos de este tipo habrá disputas sobre su aplicación marginal, que pueden comportar consecuencias prácticas graves. Son disputas usuales en derecho, donde el reto puede ser determinar si un acto dado constituye un robo, por ejemplo. Los teóricos del derecho disienten sobre la naturaleza exacta de las disputas de ese tipo y sobre cuál es la manera adecuada de darles respuesta: los llamados realistas aceptan un papel mayor y más explícito

* «En inglés, no estamos usando el presente continuo para referirnos a una costumbre o práctica». Afirmación, esta misma, que se refiere a una costumbre o práctica y que el filósofo en cuestión realiza usando, precisamente, el presente continuo. *[N. del T.]*.

para las consideraciones políticas en la resolución de casos difíciles, pero todos están de acuerdo en que tiene que haber una interpretación compartida de los casos esenciales o centrales, sin la cual las disputas sobre casos difíciles serían imposibles. En cierta medida, esto también tiene que ser así dentro de las estructuras menos formales de los debates éticos que implican esos términos sustantivos.

Algunas tradiciones otorgan gran importancia a esta veta legalista del pensamiento ético, que alienta las concepciones objetivistas de la ética, pues los casos esenciales dependen de la comprensión de esos términos éticos y su aplicación a los casos difíciles, aunque es una cuestión polémica y con consecuencias éticas, está condicionada por criterios racionales de lo que es y lo que no es un símil adecuado en los casos centrales. Es lícito discutir si una extensión dada de un término respeta apropiadamente el espíritu o principio subyacente de su aplicación en los casos esenciales. En la tradición católica, los argumentos de este estilo se conocen como casuística (la connotación peyorativa del término fue una reacción merecida a los usos retorcidos de esta técnica). El problema de la casuística, si se concibe como el procedimiento básico del pensamiento ético, no es tanto su mal uso como el hecho evidente de que el repertorio de conceptos éticos sustantivos difiere de unas culturas a otras, cambia con el tiempo y está abierto a críticas. Para que, aplicada a un conjunto local dado de conceptos, la casuística sea el procedimiento central del pensamiento ético, hacen falta más explicaciones. Se tiene que mostrar que hay categorías éticas privilegiadas que no son puramente locales. Estas categorías podrían proceder de una teoría de la naturaleza humana; de esta manera, la explicación nos remite a las preocupaciones del capítulo 3. Podrían estar dadas por un mandamiento o por una revelación de origen divino; de esta forma, si no se combina con una fundamentación de la naturaleza humana, la explicación no nos conducirá más que a lo que Spinoza llamaba «el asilo de la ignorancia». Un exponente del método casuístico podría quizá recurrir simplemente a la idea de que las categorías que preferimos son las que hemos heredado. Esta posición tiene el mérito de afrontar una verdad importante, pero no podrá afrontarla *en* verdad a no ser que sea más explícita sobre cómo podrían criticarse esas categorías.

Cuando nos apartamos del uso de los términos éticos sustantivos y consideramos cosas tales como las respuestas de las personas a preguntas sobre lo que es éticamente correcto hacer en determinadas situaciones, la analogía entre la habilidad de dar respuestas «intuitivas» (seguras

y espontáneas) a estas preguntas, por un lado, y la competencia lingüística, por el otro, parece mucho más insignificante. La habilidad para dar respuestas éticas exige algún tipo de explicación. Los casos presentados no son exactamente como los anteriores, y quien responde tiene que haber interiorizado algo que le permita responder a los casos nuevos. Pero no resulta evidente qué pueda ser. En particular, no es evidente que tenga que ser un principio, en el sentido de una descripción sumaria y fijada por medio del discurso que no descanse excesivamente en referencias a una gradación vaga («demasiado», «lo compensa», «no presta bastante atención a...»). La filosofía del lenguaje debate en qué medida la competencia lingüística, especialmente en lo semántico, puede ser plasmada en algún conjunto de reglas enunciables. En el caso de la ética, en tanto que el problema se concibe como un problema explicativo acerca de la representación de la habilidad de las personas para hacer juicios sobre casos nuevos, no necesitamos presuponer que esta capacidad descansa en una regla discursiva clara. Aristóteles pensaba que no había tal regla y que se trataba esencialmente de un tipo de juicio no explícito, de una habilidad, compartida por personas con una educación similar, para ver ciertos casos como parecidos a otros.

Esto es lo que los seguidores de Wittgenstein están dispuestos a creer sobre todo aprendizaje humano. En algún nivel último, tienen que estar en lo cierto: comprender una regla discursiva sintética implicaría compartir una apreciación de semejanzas. Pero esta concepción de la habilidad de llegar a juicios éticos compartidos (y lo mismo vale para otros tipos de juicio práctico) va más allá. No es solo que la habilidad de usar el lenguaje requiera una capacidad compartida de ver semejanzas, sino que la capacidad de ver semejanzas éticas va más allá de todo lo que puede expresarse adecuadamente mediante el lenguaje. Esto es indudablemente cierto, y es lo que los wittgensteinianos predecirían. No significa, sin embargo (y los wittgensteinianos no siempre son muy claros al respecto), que no haya ninguna explicación, a ningún nivel, de esas disposiciones humanas. Lo único que significa es que la explicación no pasa por postular la existencia de una regla enunciable, que quien se interroga ha interiorizado y consulta inconscientemente. En la medida en que nos preocupa, a un nivel explicativo, la habilidad para responder a casos nuevos, no cabe esperar que necesariamente tengamos que descubrir una regla subyacente a esa habilidad.

La analogía entre las intuiciones éticas y las lingüísticas parece muy débil habida cuenta del conflicto de intuiciones. Cuando en lingüística

surge un conflicto entre las intuiciones de dos personas, reconocemos que estamos ante dos dialectos diferentes (aunque solo difieran trivialmente); si una persona tiene intuiciones en conflicto, esto pone de manifiesto una incertidumbre que puede plantearse porque la respuesta al caso concreto está infradeterminada por el lenguaje, o quizá porque el hablante se ha educado en dos dialectos. En ninguno de estos casos, la teoría del lenguaje está obligada a *resolver* el conflicto. La teoría lingüística resolverá algunos conflictos para sus propios propósitos. De hecho, lo hace gracias a una noción idealizada de intuición, puesto que la observación de la práctica, de cómo habla realmente la gente, revela muchas incoherencias debidas a las condiciones del habla, y estas se atenúan en una concepción de la intuición como respuesta reflexiva a una pregunta sobre el lenguaje. (Los lingüistas disienten sobre en qué medida este procedimiento es legítimo). Además, se puede admitir sin duda que una teoría, que ha forjado un principio sobre la solidez de ciertas intuiciones, descarte otras intuiciones contradictorias. Descartarlas es considerarlas como prácticas anómalas, por ejemplo, o como generadoras de una práctica lingüística que el léxico ha de integrar como una singularidad, sin conexión con un principio general. Estas nociones constituyen mecanismos teóricos con los que afrontar estos conflictos.

No sucede lo mismo con las intuiciones éticas. Estas dependen, en gran parte, de la perspectiva que hay que adoptar, y una perspectiva ética idiosincrática no será abandonada si atañe a cualquier cuestión de importancia o concierne a los intereses de los demás. Aquí el objetivo de la teoría no es solo, ni siquiera en primer lugar, comprender el conflicto. Tenemos otros medios, históricos y sociológicos, para comprenderlo. El objetivo de la teoría es, más bien, resolverlo, en el sentido más radical, proporcionando una razón convincente para aceptar una intuición y no otra. La cuestión que tenemos que examinar es la de cómo una teoría ética puede tener autoridad para hacer eso.

Hay una respuesta a la pregunta que, desde ciertos presupuestos muy fuertes, tiene sentido. Asumamos que hay personas que, ante todo, están decididas a alcanzar acuerdos sobre temas éticos importantes y, de hecho, están más decididamente resueltos a alcanzar acuerdos de lo que lo están a expresar diferentes concepciones éticas del mundo. Están irreversiblemente comprometidas a vivir estrechamente vinculadas en una

sociedad[4]. Además, están resueltas a alcanzar un acuerdo, y no estarían satisfechas con el mero predominio de un conjunto de creencias. A continuación, consideran este acuerdo como una tarea que les exige llegar a principios públicamente enunciables. Y, por último, quieren que este procedimiento rija la discusión de los problemas que surjan con posterioridad a partir de los principios acordados y dirima los conflictos. En tales circunstancias, es razonable que aspiren a desplegar una teoría ética, como también lo es que usen un método que trate de salvar el mayor número posible de intuiciones y que al mismo tiempo produzca una estructura racional de principios que permitan aclarar qué intuiciones cabe abandonar o modificar. Una manera obvia de hacer esto es modificando la teoría y las intuiciones de forma recíproca hasta que encajen más o menos. Los objetivos que acabo de mencionar son, por supuesto, los de una teoría contractualista tal como la de Rawls, y el método es el que Rawls recomienda, el método de tratar de llegar a lo que él denomina un *equilibrio reflexivo* entre la teoría y las intuiciones[5].

El método es apropiado para construir una teoría ética desde esos presupuestos, pero es muy importante conocer la fortaleza de esos presupuestos. Estos implican creencias sobre cómo son las personas y cómo pueden funcionar las sociedades —*pueden* funcionar, más que funcionan, dado que es posible que ninguna sociedad haya funcionado nunca como requiere la teoría, pero es esencial para la empresa que eso sea siquiera posible—. Los presupuestos también implican ciertos ideales, y este tipo de teoría, en consecuencia, parte de dentro del mundo ético por partida doble. No solo parte del presupuesto de que los afectados aspiran a ser parte de un mundo ético, sino que también alberga la aspiración de ser parte de un tipo de mundo ético y no de otros. (Lo vimos en el capítulo 5, en la actitud que adoptan implícitamente hacia la esclavitud las partes contratantes de Rawls).

Lo fáctico y lo ideal se relacionan de manera interesante en esos presupuestos. Por un lado, hay presupuestos que se aplican a cualquier

---

[4] Este es un presupuesto tan general como confuso. Cabe destacar que desempeña una función significativa en el modelo que desarrolla Bruce Ackerman en su libro *Social Justice in the Liberal State* (New Haven, Yale University Press, 1980) [*La justicia social en el Estado liberal,* trad. de L. Rodríguez Abascal (Madrid, CEC, 1993)]; para una crítica, véase la serie de comentarios al libro de Ackerman publicados en *Ethics,* 93 (1983).

[5] En *Teoría de la justicia,* describe el método en las págs. 20 y ss., y lo refiere a Aristóteles en las págs. 48-51 [1.ª ed. inglesa; págs. 38 y ss. y 68-71 de la ed. castellana].

sociedad; por el otro, hay ideales referidos a una sociedad mejor o más racional. Entremedio, hay un área significativa de condiciones, aunque no claramente definida, que se aplican a un determinado tipo de sociedad; dicho sucintamente, a una sociedad moderna, entendida en cierta medida como una concepción ética y no solo histórica.

El primer presupuesto del procedimiento de Rawls implica un elemento que tiene que aplicarse a toda sociedad, a saber, que ha de disponer de cierto grado de homogeneidad en las creencias y de ciertas vías para resolver conflictos que tengan autoridad y eviten la violencia. Sin embargo, va mucho más allá de ser una condición mínima, en tanto que se aspira a una sociedad en la que estas vías consistan esencialmente en el debate consensual de manera abrumadora. Esta no es una necesidad social, y aquí la cuestión no es saber si las sociedades son así, sino si alguna podría serlo. A este respecto, el modelo es *liberal*. Además, es *racionalista,* en tanto que requiere que la sociedad plasme sus valores en un conjunto de principios enunciables. Una vez más, cabe decir que no es necesario que una sociedad plasme sus valores de esa manera, y está de nuevo en cuestión si podría hacerlo. Esta aspiración va un paso más allá de los presupuestos mínimos del liberalismo. Implica no solo la aspiración al debate consensual, sino también una concepción racionalista del debate consensual, una concepción característica de las sociedades modernas, o al menos de sus representaciones filosóficas y sociológicas.

Dar un paso más aquí, con un desarrollo que desborda lo que podríamos denominar racionalismo *expositivo,* implica insistir en que los conflictos entre los principios de la sociedad habrían de resolverse mediante el mismo tipo de procedimiento. Habiendo dado el primer paso racionalista, uno podría descansar en principios que se enunciaran discursivamente pero sin un orden sistemático. El requisito siguiente es el establecimiento de un procedimiento de decisión racionalista, de un método de resolución de conflictos susceptible de ser expuesto discursivamente. Es este requisito el que produce una teoría ética en el sentido pleno. Con ello se rechaza un planteamiento que también ha sido denominado, en un sentido diferente del considerado antes, «intuicionismo»; un planteamiento que propone un conjunto de principios o de consideraciones éticas pero permite que sus conflictos mutuos, al menos a partir de cierto punto, solo puedan resolverse mediante el ejercicio del juicio en cada caso concreto. Muchos teóricos éticos creen que es la misma razón la que plantea la exigencia de ir más allá del intuicio-

nismo para establecer una teoría ética mejor articulada. Piensan que lo que llamé racionalismo se sigue sin más de ser racional.

Un elemento importante de estas actividades, y en la teoría ética cuando responde a este tipo de motivación, es la aspiración a una sociedad *transparente,* en el sentido de que el funcionamiento de sus instituciones éticas no dependa de que los miembros de la comunidad no comprendan bien su funcionamiento[6]. Esta exigencia, que Rawls adopta explícitamente, encaja bien con el contractualismo liberal, pero tiene también un alcance más general, pues establece una distinción no tanto entre liberales y no liberales como entre todo aquel que mantiene esperanzas más radicales heredadas de la Ilustración y quienes no las mantienen. Muchas teorías marxistas son una representación de esto, al aspirar a una sociedad que pueda funcionar sin falsa consciencia. Esta concepción puede contribuir a conformar una crítica radical. Puede resultar útil, por ejemplo, cuestionarse en qué medida las relaciones aceptadas entre los sexos dependen de una ignorancia y una incomprensión impuestas. Los críticos radicales a veces insisten en esta cuestión siguiendo una circularidad irritante, al denunciar una falsa consciencia que (por lo que parece) consiste en no aceptar su ideología. Una utilización más eficaz de la crítica aportará otras razones para decir que se trata de una ignorancia impuesta; decir que es ignorancia, y decir que es impuesta, será hasta cierto punto independiente de la aceptación de la ideología del crítico. Esta es una condición difícil de satisfacer en muchos casos interesantes, pero hay también casos en los que no presenta ninguna complicación; la falsedad de la mala organización social es mera falsedad —mentiras, patrañas, lenguaje confuso—, al igual que su crueldad y su brutalidad no suelen ser más que crueldad y brutalidad.

La aspiración de transparencia adopta varias formas, algunas más ambiciosas que otras. No trataré de discutir aquí los interrogantes que plantean[7]. Es significativo que, si bien la transparencia es un aliado

---

[6] Rawls lo llama la condición del «carácter público»; véase *Teoría de la justicia,* pág. 133 [1.ª ed. inglesa; págs. 206-207 de la ed. castellana] y *passim,* esp. sec. 29. Para críticas en términos de falsa consciencia, véanse las referencias a la teoría crítica, capítulo 9, nota 11.

[7] El modelo defendido por Jürgen Habermas de relaciones humanas que están libres de dominación se basa en presupuestos fuertemente kantianos, y adolece también de exigir que todo se desarrolle de un modo desmedidamente explícito.

natural del liberalismo, está lejos de implicar el racionalismo. Una aspiración es la de que las relaciones sociales y éticas no han de descansar esencialmente en la ignorancia y la incomprensión de lo que son estas relaciones, y otra muy distinta la de que todas las creencias y los principios que estas implican hayan de explicitarse. Que estas son dos cuestiones diferentes es obvio en el caso de las relaciones personales, en las que es aceptable esperar que estas no descansen en el engaño y el error, pero es una estupidez creer que sus fundamentos pueden explicitarse por completo.

La empresa contractualista es coherente si adopta este enfoque y estos presupuestos, como también lo es su uso de las intuiciones para intentar alcanzar un equilibrio reflexivo. Pero todo esto no resuelve las grandes cuestiones sobre cómo se relacionan esos presupuestos con la realidad, sobre cómo puede permanecer unida una sociedad con ese grado de incomprensión ética autoconsciente, ni para qué sociedades reales representa una aspiración razonable. (Tengo la esperanza de que las ciencias sociales se planteen estas cuestiones, pero no veo claro que ninguna ciencia social haya contribuido mucho a darles respuesta).

A veces se objeta contra el método del equilibrio reflexivo que las intuiciones a las que se ajusta la teoría no representan más que creencias éticas locales, que no son necesariamente correctas. Pero, si la empresa se concibe en el modo descrito, esta no es una objeción pertinente. Se supone que las intuiciones representan nuestras creencias éticas, porque la teoría que se busca es la referida a una vida ética tal y como es *para nosotros,* y lo importante no es que las intuiciones sean en algún sentido último correctas, sino que sean las nuestras. Además, no solo son nuestras las intuiciones que moldean y son moldeadas por la teoría; también lo son las aspiraciones que, en primer lugar, motivan la teoría.

Con todo, aunque se concedan los presupuestos anteriores, el método tiene todavía un problema, y este problema puede ser lo que tiene en mente quien se pregunta si las intuiciones que suscribe el método son «realmente correctas». El problema es el de quiénes *somos*. Al describir la teoría me he referido a «una sociedad», lo que tiene un significado bastante claro cuando pensamos en una teoría contractualista, que es, en primer lugar, una teoría de la justicia social. Pero los intereses éticos desbordan estos límites, como Rawls no puede más que reconocer, y cuando nuestra atención va más allá de un orden político posible cualquier concepción específica de una sociedad se desvanece; la teoría tiende hacia esa «circunscripción natural», como la denominé, que in-

cluye a todos aquellos que podrían ser sujetos de acuerdo ético: a cualquier agente moral, como diría la teoría. Volvemos, así, a los intereses universalistas, kantianos, iniciales de una teoría de este tipo.

En cierto sentido, estos intereses encajan bien con el espíritu de la empresa contractualista y con los presupuestos que acabo de describir. Sus aspiraciones liberales y racionalistas trazan sus límites externos para incluir, además de a aquellos con quienes tenemos que alcanzar efectivamente un acuerdo, a aquellos con quienes sería éticamente deseable llegar a un acuerdo y, en último término, con cualquiera que pueda existir y sea concebible establecer acuerdos con él: la clase de los agentes racionales, los ciudadanos de la república nocional. Pero en el transcurso de este procedimiento suceden dos cosas que están conectadas. Paulatinamente, resulta menos apropiado que nos basemos en las intuiciones que pertenecen distintivamente al *nosotros* local, pues la teoría ha de ser ahora una teoría que concierne a un *nosotros* que incluye agentes que están muy alejados de nuestras costumbres locales. Por la misma razón, tenemos paulatinamente menos en lo que apoyarnos para elaborar el contenido posible de una teoría ética. Según la fórmula de Scanlon (citada en el capítulo 5), la base sobre la que un agente podría decidir «rechazar de manera razonable» un conjunto de reglas es cada vez más liviana. En último término, no hay nada que ofrecer en respuesta a esa pregunta, excepto, de nuevo, la concepción de lo que cualquier agente racional como tal tendría que rechazar, y esta concepción es muy indeterminada. Lo que la «libertad» del capítulo 4 podría proporcionarnos más específicamente, teniendo en cuenta las inmensas diferencias posibles en nuestra constitución corporal, en nuestros sentimientos y, en general, en lo que los wittgensteinianos llaman «forma de vida», es una cuestión desconcertante.

Es tan desconcertante que uno tiene que preguntarse si es la pregunta correcta. Si el modelo es el de la coexistencia con criaturas muy diferentes a nosotros, ¿por qué habría de llevarnos a imaginar una república universal en lugar de una confederación o —menos que eso, y mucho más apropiado— un mero tratado de no agresión? El código más esquemático contra las interferencias y la destrucción mutua puede ser suficiente para las partes que solo comparten la necesidad de vivir, no la de compartir una vida. Si ese código esquemático tuviese que proporcionar toda la sustancia ética de la vida en común, tendría poco que ofrecer: la vida en común exige más de lo que ofrece un individualismo meramente defensivo. Sería inapropiado reinterpretarlo de ese

modo. Pero entonces se sigue que la perspectiva universalista no determinará el contenido de la teoría ética para un grupo dado, y se cuestionará no solo si contenidos éticos diferentes pueden no ser apropiados para grupos diferentes, sino también en qué medida los presupuestos liberales y racionalistas de la teoría contractualista son de por sí apropiados para determinar el contenido de la vida ética de cada grupo.

Este fue inicialmente un problema hegeliano. Hegel criticó admirablemente el carácter «abstracto» de la moral kantiana y lo contrastó con la noción de *Sittlichkeit,* una existencia ética concretamente determinada que se expresa en las costumbres locales, una forma de vida que tiene un sentido particular para las personas que la viven. Esta concepción plantea inevitablemente la cuestión de en qué medida estas costumbres pueden continuar siendo meramente locales, y si no pueden ser criticadas, regladas o trascendidas. Para responder a esas cuestiones, Hegel apela a una concepción teleológica de la historia que implica el desarrollo de la autoconsciencia. Pocos confiarán, hoy en día, en esto; solo los marxistas más extravagantemente optimistas que todavía aceptan una interpretación materialista de la historia. Pero el problema hegeliano es el auténtico problema, por lo menos en la medida en que se pregunta cómo puede difundirse una forma de vida experimentada de modo concreto, en vez de considerar cómo puede aplicarse un programa universal. Además, las concepciones de la autoconsciencia siguen siendo muy relevantes para el problema[8].

La teoría ética contractualista no puede proporcionar el método básico de comprensión de la ética, porque ella misma necesita ser comprendida. Si se aplica demasiado ampliamente, no da resultado alguno, o no el suficiente; si se aplica demasiado restringidamente, plantea de forma insistente la cuestión de las condiciones específicas requeridas para hacerla adecuada. Puede haber circunstancias en las que las aspiraciones a un mundo mejor se expresen con más eficacia mediante el proyecto de pensar la vida ética en términos de una teoría contractualista; pero, si las hay, la idea general de una teoría semejante no basta para decirnos cuáles son.

---

[8] Véase, por ejemplo, la *Fenomenología del Espíritu,* sección «La realización de la autoconsciencia racional mediante sí misma». Véanse los comentarios de Charles Taylor, *Hegel* (Nueva York, Cambridge University Press, 1975) [*Hegel,* trad. de F. Castro Merrifield (Barcelona, Anthropos, 2010)], y Judith N. Shklar, *Freedom and Independence* (Nueva York, Cambridge University Press, 1976).

Dejando a un lado las complejidades del contractualismo, nos encontramos en otra dirección con el utilitarismo, el cual es aparentemente tan sencillo que, desde este ángulo, puede parecer atractivo. El utilitarismo no tiene ningún problema con respecto a cuánto puede alejarse de las costumbres establecidas, dado que, desde el principio, se ha alejado lo máximo posible. Por consiguiente, no tiene ningún problema sobre qué habríamos de tener en cuenta: ya nos ha dicho lo único que hemos de tener en cuenta (incluso si resulta que hay más letra pequeña de la esperada). Además, no se preocupa, al menos en este nivel amplio y sencillo, por el ajuste reflexivo con las intuiciones. Deja las intuiciones de lado.

Sin embargo, si, a este nivel, el utilitarismo no malgasta el tiempo prestando atención a nuestras intuiciones más específicas, no ha de pensarse que prescinde de toda intuición ética. Se apoya en dos. Una de ellas la caracterizó bien Henry Sidgwick en su denso y argumentado libro, *Los métodos de la ética,* al señalar que el utilitarismo requiere al menos una intuición:

> Obtengo el principio autoevidente de que el bien de un individuo no es más importante, desde el punto de vista (si puedo decirlo así) del Universo, que el de cualquier otro ... y me resulta evidente que, en tanto que ser racional, estoy destinado a aspirar al bien general —en cuanto que es accesible por medio de mis esfuerzos—, no solo a una parte específica de él[9].

El principio autoevidente de Sidgwick se presenta como un principio de racionalidad práctica; Hare expresa un principio similar en forma lingüística. Pero parece claro que lo que está en cuestión es un principio ético, y (dejando de lado un rechazo puramente egoísta de él) el desacuerdo será un desacuerdo ético. Pero el utilitarismo necesita también otro principio: el de que no hay otras consideraciones éticas fundamentales más allá de esa primera. Esta es también una intuición éti-

---

[9] El libro de Sidgwick fue publicado originalmente en 1874 y tuvo diversas reediciones con muchas modificaciones. La cita procede de la séptima edición *(The Methods of Ethics,* Londres, 1907; reimpreso en 1962), pág. 382; la frase «el punto de vista del Universo» reaparece en la pág. 420, sin la justificación. He escrito sobre Sidgwick y los problemas característicos del tipo de teoría que propone en «The Point of View of the Universe: Sidgwick and the Ambitions of Ethics», *Cambridge Review,* 7 (1982). Parte del contenido de este capítulo procede de ese artículo.

ca, y el desacuerdo que, sin duda, suscitará en muchas personas será, una vez más, un desacuerdo ético.

Se puede arrojar cierta luz sobre el segundo principio de tal manera que no parezca un presupuesto ético, sino un requisito de la racionalidad teórica: como el principio de simplicidad, según el cual los presupuestos han de ser lo más económicos posible. Pero este punto de vista comienza malinterpretando el principio de simplicidad mismo. Es una buena idea establecer un presupuesto mínimo que ofrezca la explicación más económica, pero esto no es necesariamente lo mismo que presuponer lo mínimo. El conjunto de presupuestos más efectivo no tiene por qué ser el más restringido. Un error de este tipo forma parte de la historia del empirismo, que parte del presupuesto de que la mente, antes de cualquier experiencia, está vacía. Esto presupone un contenido mínimo de la mente, pero con respecto a lo que hace falta explicar, no es un presupuesto mínimo, sino que comporta explicaciones elaboradas y muy implausibles de la evolución y del aprendizaje humanos. De modo similar, el hecho de que el equipaje con el que parte el utilitarismo sea tan ligero no proporciona *en absoluto* un presupuesto a su favor. La cuestión solo puede ser la de si su equipaje es suficiente para el viaje que tiene que hacer.

En cualquier caso, hay una pregunta fundamental sobre el viaje que todavía no ha sido respondida. ¿Por qué habría de ser apropiada la simplicidad teórica y sus criterios? Que lo sea tiene que depender de cuál sea la meta de una teoría ética. Sidgwick, por su parte, a veces parece no tener más objetivo que proyectar «la moral del sentido común en una forma científica», y resulta revelador al respecto que se preguntara que «si no tenemos que sistematizar las actividades humanas tomando la felicidad universal como su fin común, ¿sobre qué principios vamos a sistematizarlas?»[10]. Algunos teóricos posteriores también han asumido sin cuestionárselo que el sistema ético ha de intentar tener las mismas virtudes que una teoría científica.

Volveré sobre esa cuestión. En primer lugar, tenemos que examinar con más detalle cuál puede ser la relación de una teoría, en concreto, de la teoría utilitarista, con la «moral del sentido común». Para los teóricos utilitaristas, es importante que la sistematización cotidiana de las actitudes y disposiciones no signifique necesariamente que hayan de ser

---

[10] *The Methods of Ethics*, págs. 338, 406.

reemplazadas. La teoría puede justificar a veces esas actitudes; además, en ocasiones puede hacer esto incluso si las actitudes no son de por sí de tipo utilitarista. Sidgwick advirtió que tenía que ser una cuestión empírica qué motivaciones conducen al mayor bien; en particular, si es probable que la práctica de pensar en el mayor bien conduzca al mayor bien. La misma consciencia utilitarista constituye un elemento sobre el que tiene que reflexionar, y Sidgwick llegó a la conclusión de que en muchos ámbitos de la vida este proceso no ha de alentarse demasiado. Él esperaba salvar el utilitarismo de la vieja acusación de que conducía a una negación de todos los afectos naturales y a la asfixia del impulso y la espontaneidad en aras de un espíritu calculador orientado al bien universal. (Esta acusación, más que por ningún otro, fue propuesta originalmente por Godwin con su rechazo ferozmente racional a respetar ninguna consideración que un ser humano normal encontrara imperiosa). Sigdwick ofreció una explicación utilitarista de muchas disposiciones que usualmente se piensa que tienen un valor intrínseco o no utilitario. Los valores de la justicia, la honestidad, el afecto espontáneo, la lealtad a los amigos, el apego especial por los hijos, etc., podría parecer que implican un punto de vista que un utilitarista redomado no suscribiría. Pero Sidgwick insistía en que uno tiene que considerar el valor utilitarista de esos valores en función del valor del estado de cosas en el que las personas tienen esos valores. Si se hace, la justificación utilitarista irá mucho más allá de lo que se suponía.

El utilitarismo de Hare es también indirecto. Los procedimientos de ampliación de la identificación simpática no son, para él, una característica de todo pensamiento moral, sino solo de lo que llama el pensamiento «crítico». La mayor parte del tiempo pensamos a un nivel diferente, «intuitivo», sin interrogarnos sobre los efectos complejos de las acciones, sino remitiéndonos a los principios esquemáticos y sencillos que adquirimos en la infancia. No se trata meramente de reglas básicas, que pueden dejarse de lado en casos concretos en los que nos parezca apropiado, sino de principios firmemente interiorizados, de los que solo nos apartaríamos con la «mayor repugnancia», a la vez que su violación por los demás nos produce «la mayor indignación»[11]. La razón fundamental en este segundo nivel del pensamiento moral es que

[11] Estas palabras aparecen en su artículo, «Ethical Theory and Utilitarianism», en H. D. Lewis (ed.), *Contemporary British Philosophy* (Atlantic Highlands, N. J., Humanities Press, 1976); reimpreso en Sen y Williams, *Utilitarianism.*

no estamos en posición de hacer cálculos elaborados del pensamiento crítico en las situaciones cotidianas y, si tratáramos de hacerlo, es más probable que nos equivoquemos, particularmente a causa de sesgos a nuestro propio favor. Estos hechos, así como el valor consecuente del pensamiento intuitivo, pueden reconocerse gracias al pensamiento crítico. Este último, utilitarista de por sí, puede llegar a la conclusión, como hizo Sidgwick, de que la mayoría de las veces no se maximiza la utilidad pensando como un utilitarista.

Estos estilos de utilitarismo indirecto implican una concepción específica de las disposiciones que se ejercen diaria o intuitivamente; y esto plantea una cuestión importante: ¿hay algún lugar en la mente o en la sociedad en el que una teoría de este tipo pueda ubicarse coherente o aceptablemente? La teoría asigna un valor para estas disposiciones, pero es aún un valor instrumental. Las disposiciones son concebidas como medios para generar ciertas acciones, y esas acciones son los medios por los que determinados estados de cosas, que procuran el mayor bienestar, se hacen realidad. Así es cómo se ven esas disposiciones desde fuera, desde el punto de vista de la consciencia utilitarista. Pero no es lo que parecen desde dentro. De hecho, el argumento utilitarista implica que, vistas desde dentro, *no* han de entenderse así. Las disposiciones contribuyen a formar el carácter del agente que las posee, y realizarán la función que la teoría les otorga solo si el agente no concibe su carácter de un modo puramente instrumental, sino que contempla el mundo a través de ese carácter. Además, las disposiciones requieren que el agente conciba otras cosas de un modo no instrumental. Son disposiciones no solo a actuar, sino también a sentir y a juzgar, y se expresan precisamente mediante la adscripción de un valor intrínseco y no instrumental a cosas tales como la honestidad, la lealtad, etc.

Este tipo de teoría entraña una cesura incómoda, una dislocación, entre el mismo espíritu de la teoría y el espíritu que supuestamente justifica. Para cubrir este déficit o, más bien, para aceptarlo, se precisa la distinción entre teoría y práctica. Pero cuando nos preguntamos de quién es la teoría cuestionada, y de quién la práctica, la distinción resulta ser muy poco operativa. Para Sidgwick, como para muchos otros en el pasado, la distinción determinaba dos clases de personas: la clase de los teóricos que podrían tratar responsablemente la justificación utilitarista de las disposiciones no utilitaristas y la clase de quienes ponían en práctica irreflexivamente esas disposiciones. Esta perspectiva, que concuerda bien con los destacados vínculos colonialistas del utilitarismo,

puede denominarse «utilitarismo elitista». Este tiene algunas consecuencias chocantes que el propio Sidgwick aplicaba con un perfeccionismo masoquista. Por ejemplo, se planteaba la cuestión de hasta qué punto había que divulgarlo. Los utilitaristas ilustrados podían ser capaces de vivir con reglas «refinadas y complejas» que admitían excepciones en la práctica cotidiana, pero otros no podrían, y tratar de introducir esas reglas supondría «hacer más mal que bien al debilitar la moral comúnmente aceptada en vez de mejorar su cualidad». Así, los utilitaristas tienen que considerar seriamente cuánta publicidad han de dar «o al consejo o al ejemplo»:

> De este modo, según los principios utilitaristas, puede estar bien hacer y recomendar privadamente, en determinadas circunstancias, lo que no estaría bien preconizar públicamente; podría estar bien enseñar públicamente a un conjunto de personas lo que estaría mal enseñar a otros; es concebible que esté bien hacer de manera reservada lo que estaría mal hacer ante los ojos del mundo.

Sidgwick reconoce que es probable que estas conclusiones resulten chocantes; pero lo mejor es que a la mayoría de personas les continúen resultando chocantes. «Parece que la conclusión utilitarista sería», concluye Sidgwick con cierta reluctancia, «la de que la opinión de que el secreto puede hacer buena una acción que no lo sería en caso contrario habría de guardarse también relativamente en secreto; asimismo parece conveniente que la doctrina de que la moral es esotérica haya de seguir siendo esotérica»[12].

El utilitarismo elitista es indiferente a los valores de transparencia social que mencioné al examinar el contractualismo. Es una posición que hoy en día probablemente funciona mejor en la práctica que en la teoría explícita (algo que, presumiblemente, no tiene por qué lamentar). Las versiones actuales del utilitarismo indirecto, como la de Hare, suelen concebir la distinción entre teoría y práctica en términos psico-

---

[12] *The Methods of Ethics,* págs. 489-490. La concepción de Sidgwick es analizada por Derek Parfit en *Razones y personas* en el transcurso de un examen detallado de la naturaleza «autodestructiva» o tendencia a la «autodesaparición» de una teoría ética. Parfit hace hincapié en la cuestión de si el hecho de que una teoría ética tenga una u otra de estas propiedades muestra que no es verdadera. No estoy tan convencido como él de lo que esto significa. El tema del presente capítulo es qué género de vida, social o personal, haría falta para plasmar una teoría de este tipo.

lógicos más que sociales. Distinguen entre el *momento* de teorizar y el *momento* de la práctica, y usan la noción de «hora serena» de Joseph Butler en la que el moralista con disposición filosófica reflexiona sobre sus propios principios y práctica[13]. Esta versión afronta dificultades igual de grandes. No es natural suponer que un compromiso pleno con, por ejemplo, los valores de la amistad puede alternarse sin más, según un horario prescrito por la calma o la actividad, con un conjunto ajeno de reflexiones. Además, puesto que las reflexiones son de hecho ajenas, hace falta cierto tipo de olvido voluntario —un sustituto interno de las barreras de clase en las que se apoya Sidgwick— para evitar que las disposiciones con las que uno se compromete resulten perturbadas por la reflexión instrumental cuando son sometidas a presión.

El utilitarismo elitista tenía al menos el mérito de ser un tipo de realismo, en tanto que intentaba encontrar una ubicación social efectiva para la teoría. La situaba en un grupo concreto de personas: la élite utilitarista (si bien mantenía una idea engañosa sobre la naturaleza de estas personas). Algunas versiones del utilitarismo indirecto no consiguen proporcionar ninguna ubicación para la teoría. La consideran trascendente a la vida, existiendo en un espacio exterior a la práctica que ha de regular o justificar. En la versión psicológica, esto se manifiesta en una representación particular del momento de teorizar: el agente se abandona a sí mismo y lo contempla todo, incluidas sus disposiciones, desde el punto de vista del universo y, luego, ya de regreso, retoma la vida práctica. Pero cualquier procedimiento comparable de teorización tendría que formar parte de la vida, ser él mismo un tipo concreto de práctica. No puedo separar, a no ser por medio de una disociación artificial e ilusoria, el teórico que hay en mí del yo cuyas disposiciones están siendo teorizadas. En el caso del utilitarismo indirecto, esta disociación contribuye a enmascarar una dificultad concreta: el conflicto entre la concepción que de esas disposiciones tiene el teorizador y la concepción del mundo resultante de esas disposiciones.

Otras dificultades se plantean para toda tentativa de concebir la reflexión eticofilosófica como un salto al punto de vista universalista en

---

[13] Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, sermón XI. Lo que descubrimos en el momento de serenidad es que no podemos justificar un empeño ante nosotros mismos «hasta que estemos convencidos de que será para nuestra felicidad, o al menos que no irá en su contra». Cabe preguntarse si Butler pensó que estos hallazgos eran significativos.

búsqueda de una justificación, que, a continuación, remite a la práctica diaria; aun si se exige que la justificación sea coherente con el espíritu de lo que se justifica, como es típico del contractualismo, pero no del utilitarismo indirecto. Cualquier representación de ese tipo adopta, en cierto grado, el presupuesto platónico de que el agente reflexivo, en tanto que teórico, puede liberarse de la vida y del carácter que está examinando. La creencia de que podemos analizar críticamente nuestras disposiciones desde fuera, adoptar el punto de vista del universo, presupone que nos es posible comprender nuestras disposiciones y las de otras personas desde ese punto de vista sin admitir tácitamente una imagen del mundo que nos resulte más localmente familiar que cualquier otra posible; pero ni la psicología ni la historia de la reflexión ética nos dan pie a creer que los razonamientos teóricos propios de la hora serena pueden funcionar sin un sentido de la configuración moral del mundo, del tipo que se da en las disposiciones cotidianas.

¿Cuál es el atractivo de estas imágenes? ¿Qué nos impulsa a adoptar tales estructuras teóricas? Algunas de las motivaciones pueden ser muy metafísicas y generales; pueden proceder, por ejemplo, de la idea de que solo vemos el mundo tal y como es en realidad si lo observamos desde afuera, *sub specie aeternitatis*. Parte de la dignidad racional que se adscribe al intento de concebir el mundo ético desde ese punto de vista deriva de una concepción similar de la objetividad. En filosofía, se discute sobre si esta es una concepción del mundo posible, incluso para la ciencia: si nos es posible, en algún sentido último o muy radical, distanciarnos de nuestros puntos de vista sobre el mundo con el fin de alcanzar lo que he denominado, en otro lugar, una «concepción absoluta» del mismo[14]. Pero, incluso si adoptar una concepción semejante es una ambición legítima de la ciencia (como, de hecho, defenderé más adelante), esto no la convierte en un fundamento atractivo o apropiado para nuestra consciencia ética.

Por un lado, esto se debe a las diferencias existentes entre la razón práctica y la teórica; por el otro, a que la comprensión científica del

---

[14] En mi libro *Descartes: el proyecto de investigación pura*. Para el rechazo de esta concepción, véase, por ejemplo, la posición de Richard Rorty, que examino en el capítulo 8. David Wiggins, en «Truth, Invention and the Meaning of Life» (British Academy Lecture, 1976), distingue entre una perspectiva de ese tipo y los puntos de vista apropiados para la ética; aunque su elaboración de la ética guarda más semejanzas con las ideas de John McDowell, que abordo en el capítulo 8, que con el planteamiento de este libro.

mundo no solo es completamente compatible con el reconocimiento de que no ocupamos ninguna posición especial en él, sino que también incorpora ese reconocimiento[15]. El objetivo del pensamiento ético es, sin embargo, contribuir a construir un mundo que sea nuestro mundo, en el que tengamos una vida social, cultural y personal. Esto no significa que hayamos de olvidar que el mundo natural no está diseñado como nuestro hogar, y algunas de las expresiones más contundentes de la concepción intemporal, como la de Spinoza, se han dedicado a recordarnos esta verdad y su significado. Pero ello no implica que sea la perspectiva adecuada para el pensamiento ético. Si lo fuera, supondría que habríamos de pensar nuestra vida ética no solo desde esa perspectiva, sino además usando únicamente los conceptos que esta nos hace accesibles, lo que sin duda es imposible. A grandes rasgos, esos son los conceptos de la física, y es un problema todavía no resuelto en qué medida incluso el vocabulario de la psicología puede poseer ese carácter absoluto.

Esas imágenes metafísicas pueden alentar ciertas teorías éticas. Pero el impulso teórico hunde sus raíces en el pensamiento ético mismo. Para muchos, es como si la propia razón orientara el pensamiento ético en la dirección de la teoría y la sistematización. Para comprender apropiadamente las claves de la teoría en la ética, tenemos que entender por qué eso es así. A estas alturas, la cuestión importante es la de por qué hemos de creer que la reflexión necesita de la teoría. En este punto, ya es demasiado tarde para plantear la pregunta previa «¿Por qué la reflexión?». Es demasiado tarde por lo que respecta a esta investigación, dado que se le otorgó la iniciativa a Sócrates en el capítulo 1 (o, por lo menos, se le concedió en préstamo); y también es demasiado tarde en términos de la cuestión misma, pues solo podría contestarse, sin caer en el prejuicio, no tomándola en consideración. Pero es totalmente erróneo pensar que la única alternativa a la teoría ética es rechazar la reflexión y permanecer en el prejuicio irreflexivo. La teoría y el prejuicio no son las únicas alternativas para un agente inteligente, o para la filosofía.

¿Qué tipos de reflexión sobre la vida ética estimulan la teorización de manera natural? No todos lo hacen. Hay formas de reflexión que buscan la comprensión de nuestros motivos, un acceso psicológico o

---

[15] Esto no lo contradice el «principio antrópico», que discuten algunos físicos teóricos y cosmólogos, según el cual determinadas hipótesis sobre el universo quedan excluidas por el mero hecho de que existimos y podemos observarlo.

social a nuestras prácticas éticas, y aunque ambas pueden exigir ciertos tipos de teoría, la teoría ética no está entre ellas. El problema no es tampoco que esta forma de reflexión tenga un carácter explicativo, mientras que la reflexión que exige una teoría ética tendría un carácter crítico. Gran parte de la reflexión explicativa es crítica, pues revela que determinadas prácticas o sentimientos no son lo que se supone que son. Esta es una de las formas más efectivas de reflexión crítica. La que conduce a la teoría ética es de una forma diferente, dado que busca *razones justificativas*. «No creo que pueda proponerse una sola regla moral sin que alguien tenga derecho de exigir su razón», dijo Locke[16], y esta máxima, entendida de una manera determinada, conduce de modo natural a la teorización.

Muchas personas distinguen entre acabar con la vida de un feto inmaduro y acabar con la vida de un recién nacido. No todo el mundo lo hace. Muchas personas, en concreto los católicos, identifican el aborto con el infanticidio y consideran que ambos son igual de malos. Muy pocas personas equiparan el infanticidio al aborto y consideran que ambos son permisibles. En nuestro mundo moderno, parece que estos últimos son en su mayor parte entusiastas de la teoría ética[17], junto con algunos administradores implacables de los países superpoblados, sin duda. Los padres a los que les acucia la necesidad de acabar con la vida de sus bebés no consideran, por eso mismo, que ambas cosas sean equiparables. Aunque no suele mencionarse en las discusiones sobre el aborto, pocas mujeres consideran que un aborto espontáneo o temprano puedan equipararse a tener un hijo que nace muerto o que muere al poco de nacer. Así pues, para muchas personas, la distinción entre un feto y un niño es tan sólida que supone una razón con respecto a poner fin a una vida y, para la mayoría, tiene la importancia suficiente como para establecer sobre su base actitudes distintas con respecto a la cuestión de la muerte. Este es un ejemplo de una distinción que da una razón a muchas personas. Pero hay algo más para lo que no se ha dado aún ninguna razón,

---

[16] John Locke, *Ensayo sobre el entendimiento humano*, I.ii.4. [Se cita la traducción de M. E. García (Madrid, Editora Nacional, 1980).]

[17] Michael Tooley, «Abortion and Infanticide», en Cohen, Nagel y Scanlon, *The Rights and Wrongs of Abortion*, y su libro *Abortion and Infanticide* (Oxford, Clarendon Press, 1983). [El artículo se ha publicado en castellano en J. Finnis *et al.* (eds.), *Debate sobre el aborto. Cinco ensayos de filosofía moral*, trad. de M. Millán, B. Barrera y M. L. Rodríguez Tapia (Madrid, Cátedra, 1983).]

pero para lo que, según una interpretación de la máxima de Locke, sería justo exigir a esas personas que nos la den. Se trata de su práctica de usar esa distinción como una razón. ¿Cómo la justifican?

Llegados a este punto, en el que se requiere una razón para la práctica de dar razones, el número de respuestas posibles es mucho menor. Hay pocas consideraciones susceptibles de proporcionar una respuesta, y empiezan a parecerse ya a los tipos de consideración (el bienestar, el acuerdo contractual posible) que pueden figurar en una teoría ética. O, si las razones, en este nivel, aún no son de este tipo, entonces lo serán en el siguiente nivel, porque sin duda el proceso continúa. Al final, si seguimos con esta búsqueda lineal de razones, tendrá que haber al menos una práctica de dar razones para la que no se dé ninguna razón y que se sostenga por sí sola[18]. Visto de cierta manera, este resultado puede alentar el principio de simplificación que mencioné anteriormente: si bien es irracional suscribir un principio no racionalizado, también es bueno ser tan poco irracional como sea posible. Otras personas, con un temple diferente, se preguntarán por qué, si al final vamos a tener que ir a parar a una práctica de dar razones injustificada, no podemos quedarnos con varias de estas prácticas. Una vez que nos damos cuenta de que resulta imposible racionalizarlo todo, no es necesario pensar que la mejor alternativa es el proyecto de racionalizar todo lo que sea posible. Podríamos concluir, en cambio, que estamos yendo en la dirección equivocada.

Sin embargo, no tiene mucho sentido discutir qué actitud cabe tomar hacia los resultados de este modelo lineal, dado que el modelo mismo es erróneo. Ningún procedimiento de dar razones satisface esta concepción ni en las ciencias ni en ningún otro sitio. En cuestiones teóricas, la empresa fundamentista de hacer descansar la estructura del conocimiento en cierta clase privilegiada de afirmaciones, en general, ha sido desplazada a favor de un modelo de tipo holista, en el que ciertas creencias pueden cuestionarse, justificarse o ajustarse, siempre que las demás se mantengan constantes, pero no hay ningún proceso por el cual puedan cuestionarse, o justificarse, todas a la vez, en términos de (casi) nada. Según la famosa metáfora de Neurath, reparamos el barco sin dejar navegar en él.

---

[18] Por supuesto, hay otra posibilidad, desalentadora, que ha surgido ya en el debate sobre el utilitarismo indirecto: que el principio, aplicado a sí mismo en la práctica, dé razones contra sí mismo.

Si renunciamos al modelo lineal, podremos tener aún la posibilidad de disponer de una razón para cada práctica; lo que habremos perdido es la posibilidad de que haya una única razón para todo. No obstante, en el caso de la ética, incluso el requisito más endeble, que haya una razón u otra para cada práctica, tendrá que entenderse de un modo muy poco exigente si queremos que sea satisfecho. Aunque seamos capaces de mostrar cómo una práctica dada encaja con otras prácticas de un modo que tenga sentido social y psicológico, podemos no ser capaces de hallar nada que satisfaga la exigencia de justificación que hace alguien que esté fuera de esas prácticas, ni siquiera de justificárnoslas a nosotros mismos en un sentido genuino. Una práctica puede estar tan directamente conectada con nuestra experiencia que la razón que nos proporcione la consideraremos más fuerte que cualquier otra de las razones que pudiéramos dar a su favor.

Esto queda de manifiesto en los intentos de algunos teóricos de reemplazar las categorías que nos proporcionan razones por otras que se supone que son más sistemáticas. Así Michael Tooley, el teórico que quiere que nos acostumbremos a la idea del infanticidio, reclama que la categoría de *persona* sea la noción operativa en ese tipo de cuestiones: determinados deberes son exclusivos de las personas; la infancia está excluida de la clase privilegiada en un extremo, como lo está la demencia senil en el otro. Esta sugerente propuesta descansa en un concepto engañoso: la categoría de persona, aunque haya sido muy fructífera para determinada filosofía moral, es un fundamento insuficiente para el pensamiento ético, concretamente, porque aparenta ser una noción sortal o clasificatoria, cuando en realidad lo que hace es señalar características que se manifiestan casi siempre en grados: la responsabilidad, la autoconsciencia, la capacidad reflexiva, etc. Parece como si tratáramos con una clase o tipo determinado de criatura, cuando en realidad estamos considerando de forma vaga a los seres humanos que obtienen un promedio sobre una escala[19]. Para empeorar aún las cosas, el promedio que es adecuado para ciertos objetivos no lo es para otros. Si la noción *persona* implica algo así como «responsabilidad moral completa», la edad mínima que tradicionalmente se ha considerado necesaria para formar parte de esa clase es siete, pero cualquiera que haya convivido

[19] Para una caracterización de «persona» que no admite el tipo de argumento que defiende Tooley, véase David Wiggins, *Sameness and Substance* (Cambridge, Mass., Harvard University Press, 1980), cap. 6, esp. págs. 169-172.

con un niño de seis años, o de dos, tiene razones poderosas para pensar que también son personas. Tooley ni siquiera los considera lo bastante pre-personas como para permitir que se les mate.

Los defectos que tiene la categoría teórica de *persona* constituyen un fallo de esa propuesta concreta, pero también ilustran los fallos del empeño general por teorizar. ¿Cómo podemos detectar la debilidad de un concepto teórico si no es por referencia a las distinciones cotidianas a las que se supone que reemplaza o justifica y gracias a un sentido de la vida que se supone que nos ayuda a dirigir? Lejos de tener una autoridad especial por pertenecer a una teoría, es probable que estas concepciones sean más arbitrarias, con respecto a lo que se les exige, que aquellas a las que se supone que reemplazan.

Si la teoría ha de ser un producto de la *razón misma,* que es la tesis que estoy considerando, no lo será porque toda razón requiera de una razón. ¿Puede ser, de alguna otra manera, un producto de la exigencia de racionalidad? Nada de lo que se ha dicho hasta el momento debería inducirnos a pensar que las distinciones tradicionales están a salvo de la crítica; las prácticas que discriminan entre diferentes grupos de personas pueden precisar, ciertamente, una justificación, si no vamos a conformarnos con las tradiciones irreflexivas que proporcionan el paradigma de los prejuicios. Estos prejuicios parecen sencillamente irracionales; la racionalidad exige su crítica, así como que se eliminen si no están justificados; su crítica exige una razón para hacer esas discriminaciones. Si este proyecto se lleva lo suficientemente lejos, ¿cómo podremos evitar la teoría ética, a fin de cuentas?

La primera cuestión que se plantea es la de en qué medida lo incorrecto de esas prácticas depende de que sean *irracionales.* No es correcto, por injusto, tratar a los negros o a las mujeres peor que a los blancos o a los hombres y, cuando las prácticas se imponen, también suele ser cruel. Es irracional, podemos decir, porque es incoherente. Las razones no se aplican por igual, y eso ofende al principio formal de universalización. Pero esto casi nunca es lo incorrecto; si hay algo que sea irracional en tales asuntos, tiene que ser otra cosa. El principio formal de universalización, que está fuera de toda disputa, es lo que describí antes como el principio de «suficiente es suficiente», que afirma que, si determinada consideración es verdaderamente una razón suficiente para realizar determinada acción en un caso concreto, lo es también en otro caso. Pero la discriminación y el prejuicio pueden funcionar sobre esa base. Alguien que considere que *ser mujer* constituye de por sí una

razón para dispensar un trato discriminatorio no comete una ofensa contra el principio «suficiente es suficiente» solo por actuar de esta manera. La comete, y es incoherente e irracional si considera la inteligencia y la fiabilidad como fundamentos supuestamente suficientes para contratar a un hombre, pero no si se trata de una mujer. Evitaría, por lo menos, esta acusación si explicitara que, en el primer caso, las razones para contratarlo eran que el candidato era inteligente, fiable y hombre. Por supuesto, no sería más justo por exponer abiertamente sus razones, pero sí sería coherente.

Hay otra manera en que esta persona puede ser irracional. Considerar que *ser hombre* constituye una razón puede no tener sentido en el contexto de lo que se está haciendo; en este caso, la contratación laboral de alguien para que haga un trabajo eficientemente. O, si se la dota de sentido mediante una comprensión restrictiva de lo que se está haciendo, tendrá menos sentido cuando vaya más allá de esta comprensión y se interrogue sobre el interés de recurrir a los prejuicios de otras personas, los cuales no son mejores que los suyos. (Puede decir que no se atreve o no puede permitirse ir contra esos prejuicios, pero se trataría entonces de una razón diferente). Este tipo de irracionalidad no se desenmascara ni se cura invocando una teoría ética, sino obligando a la persona a reflexionar sobre lo que está haciendo. Como antes, esto puede exigir un tipo diferente de comprensión teórica, y requerirá otros valores.

Analizar lo que hace realmente una persona que discrimina es más necesario, y está más alejado de la teoría ética, en el caso más cotidiano en el que la persona en cuestión no admite que la razón por la que discrimina sea que «se trata de un negro» o «se trata de una mujer». La racionalización reemplaza a la discriminación explícita, y ofrecerá una razón que puede ser relevante pero que cree solo porque se ajusta a su propósito. Esto es todavía algo profundamente irracional, pero se trata de una irracionalidad de la creencia (del autoengaño o del fraude social) más que de una irracionalidad debida a la resistencia al poderío de una teoría ética. El estudio de la irracionalidad en la práctica social ha de centrarse en estas áreas, y requiere investigaciones más detalladas y sustantivas que las consideraciones esquemáticas de la teoría filosófica.

La consecuencia principal de esta discusión para el razonamiento ético es que la crítica reflexiva ha de tomar, fundamentalmente, la dirección contraria a la incentivada por la teoría ética. Es característico de la teoría ir en busca de consideraciones que sean muy generales, con el

menor contenido distintivo posible, porque su cometido es tratar de sistematizar y porque pretende representar todas las razones posibles como aplicaciones de otras razones. Pero la reflexión crítica ha de intentar alcanzar la mayor cota de comprensión compartida que se pueda hallar en cualquier cuestión y usar cualquier material ético que, en el contexto de la discusión reflexiva, tenga algún sentido e inspire lealtad. Por supuesto, se darán cosas por sentadas, pero toda reflexión seria tiene que saber que ha de hacerlo. La única tarea seria es vivir y, tras reflexionar, tenemos que vivir. Además, aunque la distinción entre teoría y práctica nos induce a olvidarlo, también tenemos que vivir cuando estamos reflexionando. Por regla general, la teoría presupone que es probable que tengamos demasiadas ideas éticas, algunas de las cuales es fácil que resulten ser meros prejuicios. Nuestro mayor problema es que realmente no tenemos demasiadas ideas éticas, sino demasiado pocas, y tenemos que abrigar tantas como podamos.

«Prejuicio» es una palabra impactante y ambigua, cuyas relaciones con la teoría son igualmente ambiguas. Ha desempeñado un papel importante en la tradición cartesiana, en la que cualquier creencia supone un prejuicio sin fundamento. En este sentido, se opone sin duda a la teoría, pero, como ya he dicho, en este sentido todo es prejuicio, tanto en ciencia como en ética. En un sentido más restringido, «prejuicio» designa a toda creencia mantenida solo porque no se ha reflexionado sobre ella. Así, es posible que no podamos evitar los prejuicios, pero, en cualquier caso, no hace falta que la reflexión exigida, a la que sobrevivirán algunas creencias, sea la reflexión de la teoría ética. Sin embargo, cabe insistir en ello: un prejuicio de tipo racista o sexista normalmente es una creencia que está a salvo de la reflexión porque conviene a los intereses de quienes creen en ella. He mencionado ciertos tipos de irracionalidad que se pueden incluir en este tipo de creencia, y ciertos tipos de reflexión se le podrían aplicar. Mi esbozo se aplica a la situación en la que hay elementos, tales como la comprensión de los roles sociales implicados, que pueden ser objeto de reflexión por parte de un miembro de la sociedad. Hay otras cuestiones que se plantean cuando se considera una sociedad desde un punto de vista externo a ella y el prejuicio se ve como algo colectivo, inquebrantable desde dentro. Esto suscita otro tipo de reflexión, de la que nos ocuparemos en el capítulo 9, en el que examinaremos algunos problemas del relativismo.

No obstante, todas estas reflexiones, realizadas desde dentro de una sociedad o con respecto a otras, se apoyarán en la experiencia humana y

se vincularán a los intereses humanos. Si dejamos las cosas como están, dirán algunos, se tratará solo de un prejuicio más. En último término, la igual consideración ha de extenderse, más allá de la humanidad, a todo lo que sea susceptible de ser objeto de consideración. Así pues, vemos que el utilitarismo no es solo la teoría ética más ambiciosa, sino también la que va más lejos en su ambición de usar la teoría para combatir el prejuicio. Esta ambición encaja con la actitud del observador ideal y con sus requerimientos mínimos: de lo único que puede hacer uso es de la idea de bienestar, que supuestamente puede aplicarse a todo lo que merezca nuestra atención. Pero ya hemos visto que esta perspectiva nos reporta muy poco y no nos proporciona un camino de vuelta coherente que restablezca las preocupaciones de nuestra vida ética local, esto es, de nuestra vida.

La palabra «especismo»[20] se ha usado para designar una actitud que algunos juzgan que es nuestro último prejuicio: privilegiar a la humanidad. Es más revelador denominarlo «humanismo», y no se trata de un prejuicio[21]. Ver el mundo desde un punto de vista humano no es algo

---

[20] Al parecer, el término fue acuñado por Richard D. Ryder en *Victims of Science: The Use of Animals in Research* (Londres, Davis-Poynter, 1975). Hay una extensa bibliografía reciente sobre la ética de nuestras relaciones con los animales, gran parte de la cual nos insta a respetar los denominados «derechos de los animales». No puedo ocuparme adecuadamente de esta cuestión aquí, aunque cabe plantear tres asuntos de una forma muy sumaria. En primer lugar, hay buenas razones para no infligir dolor a los animales, pero no se gana nada, que no sea meramente retórico, en fundamentar esto en *derechos*. Los derechos son tipos distintivos de razones éticas, y se explican mejor en términos de dar seguridad a expectativas (véase el capítulo 10, sobre las obligaciones), una consideración que no se aplica a los demás animales. Para un punto de vista opuesto, véase Tom Regan, *The Case for Animal Rights* (Berkeley, University of California Press, 1983). En segundo lugar, si se toma como base utilitarista más simple la minimización del nivel de dolor, no está claro por qué no hemos de comprometernos, como señaló hace años Ritchie en *Natural Rights* (1984), a dedicar todo el tiempo que podamos a la vigilancia de la naturaleza. Por último, hay una línea argumental diferente, que funda nuestras relaciones con los animales en una teleología general, que nos anima a verlos no como un recurso sino como seres que comparten el mundo con nosotros; véase Stephen R. L. Clark, *The Moral Status of Animals* (Nueva York, Oxford University Press, 1977). Pero no entiendo por qué, según una concepción realista de nuestras relaciones «naturales» recíprocas con los animales, ha de excluirse que nos los comamos.

[21] Como prejuicio, no difiere del todo del mencionado por el capitán Jack Absolute, personaje de Sheridan: «Reconozco que debería elegir una esposa que tenga el número habitual de extremidades y una cantidad limitada de trasero; y aunque un único ojo puede ser muy agradable, no obstante, como el prejuicio siempre se ha pronunciado a favor de dos, no me gustaría singularizarme a este respecto» *(Los rivales,* III, I).

absurdo para los seres humanos. A veces se dice que esta concepción implica considerar que los seres humanos son las criaturas más importantes o valiosas del universo. Sin duda, esto sería absurdo, pero no lo implica. Suponer que lo hace es cometer el error de identificar el punto de vista del universo y el punto de vista humano. Nadie habría de defender esas tesis sobre la importancia de los seres humanos para el universo: lo esencial es la importancia de los seres humanos para los seres humanos.

Preocuparnos por los animales no humanos es parte de la vida humana, pero podemos adquirir esta actitud, cultivarla y enseñarla solo por lo que respecta a nuestra comprensión de nosotros mismos. Los seres humanos tienen esa comprensión y son objeto de ella a un tiempo; este es uno de los aspectos básicos en los que nuestras relaciones éticas mutuas tienen que ser siempre diferentes de nuestras relaciones con otros animales. Antes de llegar a la cuestión de decidir cómo tratar a los animales, cabe subrayar en primer lugar que la única cuestión que se plantea es: ¿cómo han de ser tratados? La elección es, únicamente, entre si los animales se benefician de nuestras prácticas o si son dañados por ellas. Es por esto que el especismo se modela falsamente sobre el racismo o el sexismo, que son realmente prejuicios. Suponer que hay una comprensión del mundo blanca o masculina ineliminable, y pensar que la única elección es si los negros o las mujeres se beneficiarán de «nuestras» prácticas (blancas, masculinas) o serán dañados por ellas, supone ser ya víctima del prejuicio. Pero, en el caso de las relaciones de los humanos con los animales, los análogos de esas ideas son realmente correctos.

Nuestros argumentos tienen que fundamentarse en el punto de vista humano; no pueden derivarse de un punto de vista que no sea el de nadie. No se trata, como pretenden las formas más robustas de teoría ética, de que la razón nos haga ir más allá de la humanidad. Las exigencias más urgentes de la humanidad son, como lo fueron siempre, que reunamos todos los recursos posibles que nos ayuden a respetarla.

Capítulo 7

El giro lingüístico

He estado examinando diversos proyectos que tratan de fundamentar el pensamiento ético. Ninguno de estos enfoques pretendía ponernos en posición de deducir conclusiones éticas mediante la derivación lógica de valores a partir de hechos, o de recomendaciones prácticas a partir de meras descripciones del mundo, o un *debe* de un *es*. El planteamiento kantiano buscaba las precondiciones para ser un agente racional que presuntamente introducía las consideraciones éticas. Las personas que rechazaran esas consideraciones estarían confundidas con respecto a sus relaciones prácticas con el mundo o bien, si se diera un conflicto lógico, no sería entre un *debe* y un *es*, sino entre varios *debe*, todos los cuales estarían comprometidas a aceptar. En cuanto a Aristóteles, su preocupación era esencialmente la de determinar qué es más razonable que tratemos de conseguir. Su manera de dar cuenta de esto se sustenta en una teoría teleológica de la naturaleza que no podemos aceptar hoy en día, pero no depende de la deducción lógica de valores a partir de otra cosa.

Hay cuestiones en filosofía moral que conectan los juicios éticos con la razón práctica de una manera menos directa de lo que lo hacen la mayoría de los temas que hemos abordado hasta el momento. ¿Puede haber conocimiento ético? Si lo hay, ¿qué relaciones guarda con el conocimiento científico? Al examinar estas cuestiones tampoco me ocuparé de la deducción lógica de valores a partir de hechos. En el capítulo

siguiente discutiré si hay una distinción entre hecho y valor, y dónde cabe situarla. Pero, de todas formas, la distinción no será primordialmente lógica ni, menos aún, la descubriremos en el uso de las palabras.

Hasta hace poco, esa manera de proceder y esas conclusiones habrían resultado sorprendentes. Los debates sobre la deducción de valores a partir de hechos, o la cuestión adyacente de cómo definir los conceptos éticos en términos no éticos, han predominado en la filosofía moral del siglo xx, a veces haciendo caso omiso de todo lo demás. Tras estos debates se encuentra una gran idea, o más de una, a propósito de las relaciones entre el valor, el conocimiento del mundo y la libertad. La filosofía reciente a menudo ha dado una representación errónea de estas cuestiones, al debatirlas como si se tratara de la definición de las palabras. Ya me he referido a esto sucintamente en una o dos ocasiones. En este capítulo, quiero examinar más en general los asuntos planteados por un método lingüístico. El debate tiene que partir de una noción que no puede ignorar ninguna posición en filosofía moral: la «falacia naturalista».

La idea de que la ética ha de prestar especial atención a las definiciones fue alentada sobremanera por Moore, quien, en *Principia Ethica* (1903), introdujo una serie de ideas sobre el bien: sería una cualidad no natural, simple, que no podría ser definida[1]. Y acusó a quienes trataban de definir el bien de cometer la falacia naturalista. Es difícil encontrar en la historia de la filosofía un nombre tan espectacularmente erróneo para una idea. Para empezar, no está claro por qué las personas criticadas estaban cometiendo una falacia (que es un fallo en la inferencia) en lugar de lo que Moore consideraba un error, o simplemente redefiniendo una palabra. Pero más importante aún es que una palabra útil como «naturalismo» se pusiera al servicio de un propósito mal concebido. Con anterioridad, una concepción naturalista de la ética se oponía a una concepción sobrenaturalista, lo que significaba que la ética había de entenderse en términos seculares, sin ninguna referencia a Dios o a cualquier otra auto-

[1] Para un examen de esta concepción, véase la sección «Bueno» de mi libro *Introducción a la ética*. En «A Reply to my Critics», en P. A. Schilpp (ed.), *The Philosophy of G. E. Moore* (La Salle, Open Court Publishing Co., 1942), Moore admitió que la explicación ofrecida en *Principia Ethica* de la distinción entre propiedades naturales y no naturales era bastante insatisfactoria, pero no la abandonó del todo.

ridad trascendente. Se refería al tipo de concepción ética que parte de la idea de que el hombre es parte de la naturaleza. El punto de vista de Aristóteles es naturalista en este sentido, como lo es el utilitarismo de Mill y la mayoría de las obras éticas modernas, incluida esta. Las concepciones naturalistas, en este sentido amplio y útil, no cometen necesariamente la «falacia naturalista». Aristóteles no lo hizo, y no veo ninguna razón por la que cabría pensar que lo hiciera Mill (con quien Moore se mostró particularmente desagradable). Lo que causa aún más confusión es que no todo aquel que, según Moore, cometía esta falacia era también naturalista en el sentido amplio y útil. Algunos de los transgresores más llamativos eran antinaturalistas en el sentido amplio y útil, que definían el bien como aquello que Dios ordena o quiere.

Esto último muestra que aquí hay algo más que un problema terminológico. Se plantea un reto teórico importante, que surgió gradualmente a partir de los debates sobre la falacia naturalista después de que Moore introdujera dicha expresión. Si la falacia constituye un error significativo, ¿qué se nos exige exactamente que evitemos? No se trata solo de la prohibición de definir *bien* en términos naturalistas (en el sentido amplio y útil); como acabamos de ver, también prohíbe las definiciones no naturalistas. Así, ¿quizá se trate solo de prohibir una definición del *bien* en términos de cualquier otra cosa? Moore mantenía esta posición, pero (como mencioné en el capítulo 1) estaba dispuesto a definir *correcto* en términos de *bien,* y era una posibilidad que otros teóricos adoptaran una de las rutas reduccionistas alternativas a la vez que preservaban la prohibición mooreana de cometer la falacia naturalista. Esto saca a la luz que la doctrina de la falacia naturalista no es, o al menos dejó pronto de ser, una prohibición relativa solo a la definición de *bien.* En realidad, se entendió que establecía la existencia de dos clases de expresiones. Una clase contenía *bien* y *correcto,* entre otros términos que se etiquetaron, por ejemplo, como términos «evaluativos». La otra clase, no evaluativa, contenía una gran variedad de elementos, como afirmaciones de hecho, verdades matemáticas, además de cosas tales como afirmaciones sobre Dios (a menos que, debido a alguna razón independiente, estas fueran evaluativas, como la afirmación «Dios es bueno»). Así pues, la falacia naturalista constituye el intento de definir todo término perteneciente a la primera clase solo por medio de términos pertenecientes a la segunda.

No es solo una cuestión de definiciones. La prohibición excluye cualquier intento de deducir una conclusión evaluativa a partir de pre-

misas enteramente no evaluativas. (Prohibir las definiciones es un caso especial de esto, dado que una definición es un tipo de equivalencia lógica o implicación bidireccional). La ampliación de esta prohibición no solo excluye los intentos de definir el *bien*, sino también algo que Hume había tachado de sospechoso, el intento de derivar el *debe* del *ser*. El paso de proposiciones que contienen *es* a otras que contienen *debe*, señaló Hume, resulta, sin embargo, de la mayor importancia.

> En efecto, en cuanto que este *debe* o *no debe* expresa alguna nueva relación o afirmación, es necesario que ésta sea observada y explicada y que al mismo tiempo se dé razón de algo que parece absolutamente inconcebible, a saber: cómo es posible que esta nueva relación se deduzca de otras completamente diferentes[2].

Se ha dudado, con razón, de si en este pasaje Hume quería decir aquello para lo que se ha usado posteriormente[3]. Él pensaba, y dijo explícitamente, que prestar atención a esta cuestión «nos permitiría ver que la distinción entre vicio y virtud ni está basada meramente en relaciones de los objetos ni es percibida por la razón». Pero no hay relación directa entre una conclusión de este tipo y los problemas relativos a la definición y la deducción lógica.

La expresión «falacia naturalista» se suele usar hoy en día para las infracciones de la prohibición de derivar el *debe* del *es*. Al mismo tiempo, retiene su fuerza original con respecto a la definición de *bien* y de otras expresiones evaluativas semejantes. Se necesita una teoría suplementaria para explicar por qué una prohibición relativa a *debe* implica una prohibición relativa a *bien*. Esa teoría adopta la forma de aplicar la estrategia reduccionista al tratar *bien* como definible en términos de *debe*. Esta es una de las motivaciones más profundas a favor de la estrategia reduccionista. Si estamos convencidos de que tiene que haber dos clases fundamentales de expresiones, una relativa al valor y la otra a los hechos, es natural concebir como básico a uno de los miembros de la clase del valor, mientras que los demás han de definirse en términos de este.

---

[2] David Hume, *Tratado de la naturaleza humana* (1793), III.i.i [trad. de Félix Duque (Madrid, Editora Nacional, 1977)].

[3] Para diversas interpretaciones de lo que quiere decir Hume, véase W. D. Hudson (ed.), *The Is-Ought Question* (Nueva York, St. Martin's Press, 1969); así como John Mackie, *Hume's Moral Theory* (Boston, Routledge and Kegan Paul, 1980), págs. 61-63, y *Ética: inventando el bien y el mal*, págs. 64-73 [ed. inglesa].

Esto parece lo más natural si consideramos un modo concreto en el que los teóricos modernos han tratado de explicar por qué no podemos derivar lo evaluativo de lo no evaluativo. El propio Moore pensaba que la bondad era indefinible debido a que es una cualidad simple no natural. También pensaba que la presencia de esta cualidad era detectada por la intuición, en el sentido de esa aptitud intelectual, examinada en el último capítulo, que era la base de un tipo de «intuicionismo». El intuicionismo, en ese sentido, no explica gran cosa ni contribuye significativamente a explicar por qué los valores no se pueden derivar de los hechos.

Obras más recientes han intentado dar una explicación mejor. Presentan como central la prohibición de derivar el *debe* del *es*. La concepción principal es el *precriptivismo,* desarrollado por Hare, que explica la función del *debe* como prescriptiva de una acción o de decir a alguien qué hacer. El *debe* se concibe de modo semejante a un imperativo: hablando estrictamente, una afirmación que emplea un *debe* usado en el modo prescriptivo normal es una expresión universal que implica imperativos que se aplican a todos los agentes en todas las circunstancias similares. (Hemos visto ya el uso de esta idea en el desarrollo de Hare del utilitarismo del Agente Mundial). Según esta interpretación, será más revelador llamar *prescriptivo* a lo que hasta ahora he llamado *evolutivo,* y es lo prescriptivo lo que no puede derivarse de manera válida de otra clase de afirmaciones; una clase que, en contraste con esta, se denomina adecuadamente *descriptivo.* La explicación de esta prohibición resulta ahora del todo evidente. Lo prescriptivo hace algo, a saber, decir a las personas que actúen de determinada manera, cosa que lo descriptivo, por sí mismo, no puede hacer. Esto proporciona claramente una razón de la prohibición, pero si la prohibición tiene que ser tan general como se esperaba inicialmente, y ha de explicar las relaciones básicas entre hecho y valor, entonces (como ya he dicho) ha de extenderse a *bien* y a otras aplicaciones del argumento original contra la falacia naturalista. Lo evaluativo en general tendrá que reducirse a lo prescriptivo, tal y como se ha explicado hasta ahora; además, la teoría resultante cubriría, como hacía la doctrina de Moore, tanto lo no ético como lo ético. Lo que cabe demostrar, entonces, es que al decir que algo es bueno o malo, admirable o despreciable, excepcional o inferior en su género, estamos en realidad diciendo a los demás o a nosotros mismos que hagan o hagamos algo (o, como afirma normalmente esta explicación, que elijamos algo). Cualquier evaluación tiene que estar conectada con la acción.

Este resultado no es del todo fácil de aceptar. Por ejemplo, parece falso respecto al espíritu de muchas evaluaciones estéticas: parece exigir que nuestra perspectiva básica sobre el valor de un cuadro sea, a grandes rasgos, la de un coleccionista potencial. Incluso dentro del ámbito ético, suponer que las personas reconocidas como buenas son personas que se nos exhorta a imitar revela una concepción demasiado limitada de los méritos humanos. Hare, al explicar la fuerza prescriptiva de la evaluación, afirma que:

> Si decimos [de un hotel determinado] que es un hotel mejor que el que está al otro lado de la calle, esto implica un sentido de «mejor que» (en el sentido prescriptivo) por el que una persona que asiente oralmente a nuestro juicio, y que, cuando tiene que elegir entre ambos hoteles (sin que varíen otros factores, como el precio) se decanta por el segundo, tiene que haber dicho algo que en realidad no piensa[4].

Creer que algo es mejor, en el sentido prescriptivo, prosigue Hare, es preferirlo. Pero esto no es exactamente lo que él piensa, como revela su ejemplo. No es, por ejemplo, preferir un hotel sobre la base del precio. Tiene que preferirse por sus méritos en tanto que hotel. Pero una vez hemos introducido la noción de los méritos de un hotel (o de cualquier otra cosa), la conexión lógica con la preferencia parece errónea, y yo no veo ninguna razón para pensar que exista ese *sentido* de «mejor que», al menos si lo que está en cuestión son los méritos de una cosa concreta de ese tipo, como un hotel. Los méritos de los hoteles no son una cosa muy definida y dejan amplio margen al gusto personal, pero incluso en ese caso, y aunque un hotel sea (en principio) algo que tiene la función de servir al gusto de los clientes, puedo distinguir entre los méritos de un hotel y lo que yo, por excelentes razones, prefiero en realidad. «Sencillamente, no me gusta alojarme en buenos hoteles» es una afirmación del todo inteligible.

Esto saca a relucir la debilidad básica de las teorías prescriptivistas de lo evaluativo. Cuando evaluamos cierta cosa, hay determinados criterios relativos a los méritos de las cosas de ese tipo, incluso si en un caso particular pueden tener un alcance solo local, o ser vagos o indeterminados. Aunque hay muchas elecciones que estarán adecuadamente determinadas por los méritos de la cosa, porque la persona que elige

---

[4] Hare, *Moral Thinking*, pág. 21.

busca una cosa buena de ese tipo, siempre subsiste la posibilidad de que la elección de una persona dada no se rija directamente por cómo la persona ve los méritos de esa cosa. Para muchos tipos de cosas, podemos distinguir entre pensar que un elemento dado es bueno en su tipo y que realmente nos guste, lo queramos o lo elijamos; además, nuestra habilidad para hacer distinciones muestra que comprendemos que los méritos de la cosa en cuestión pueden ir más allá de nuestros intereses o de nuestra capacidad de respuesta. La filosofía no puede hacer lógicamente obligatoria la actitud que tenía un hombre que conocí que, en una de esas discusiones sobre qué música mala nos distrae más, dijo que «Me parece que puedo sobrevivir con una dieta de obras maestras».

Hay, pues, problemas serios relativos a la función que puede desempeñar la distinción entre *es* y *debe*. Este es un asunto importante ligado a la presentación lingüística de la distinción entre hechos y valores. Las relaciones entre *es* y *debe* proporcionan uno de los pocos casos en los que se ha defendido claramente una distinción de este tipo en términos lingüísticos: haría falta que esta se generalizara al resto del lenguaje evaluativo para que los resultados fueran tan significativos como se supone que es la distinción.

Más allá de esto, también hace falta que nos preguntemos qué puede establecerse exactamente respecto a la relación entre el *es* y el *debe*. Se ha de poder encontrar una verdad evidente en la distinción *es/debe*, aunque puede que esa no sea la mejor manera de expresarlo. Es una cuestión relativa al razonamiento práctico. Es verdad que el *debe* no puede deducirse del *es*, si el *debe* se considera equivalente al *he de* que aparece en la pregunta práctica «¿Qué he de hacer?» y en la respuesta a ella «considerando todos los factores». Tal respuesta, la conclusión de un razonamiento práctico, no puede deducirse lógicamente de las premisas que la sostienen[5]. Esto resulta obvio si no incluimos entre las premisas afirmaciones sobre lo que queremos: lo que encontramos más razonable hacer tiene que depender de lo que queremos. Pero, si introducimos todos los enunciados relevantes sobre lo que queremos, incluso si introducimos los principios generales de la teoría de la decisión que estamos usando (que, en alguna medida, tienen que ser una cuestión de elección o de temperamento), no será aún un

---

[5] Estoy en deuda con David Wiggins en este punto; véase «Truth, Invention, and the Meaning of Life» (British Academy Lecture, 1976), y «Deliberation and Practical Reason», en Amelie O. Rorty (ed.), *Essays on Aristotle's Ethics* [«La deliberación y la razón práctica», trad. de J. J. Utrilla, en Joseph Raz (comp.) *Razonamiento práctico* (México, FCE, 1986)].

asunto de deducción lógica llegar a una conclusión sobre lo que hemos de hacer considerando todos los factores. Esto siempre exige que determinemos qué es lo que, en esta ocasión particular, y a la vista de todo lo relevante, juzgamos más importante.

En muchas ocasiones resultará completamente evidente qué es lo más importante y «determinarlo» no requerirá ningún momento de decisión, pero resta aún un paso ulterior más allá del *input* recibido. Si trato de incluir este paso en el *input,* por ejemplo, diciendo «considero que los factores ABC son los más importantes», entonces, o bien esta es una precisión sobre mis disposiciones («ese es el tipo de persona que soy»), lo que implica una determinación suplementaria; o bien expresa lo que se determina en esta ocasión, y la afirmación no es en realidad parte del *input,* sino una anticipación de la conclusión.

Incluso a partir de esta conclusión «completa», hay aún un paso más que dar para llegar a la acción, y puede abrirse una brecha entre esta conclusión y la intención de actuar. Este es el ámbito de la acrasia. Pero la eventualidad de la existencia de esa brecha no da pie de nuevo, por así decir, a que la conclusión haya de estar lógicamente implicada por el *input*. Es un error pensar que todo se circunscribe al razonamiento lógico hasta que llegamos a la decisión de actuar. El problema de la acrasia afecta a las relaciones entre la intención y nuestro mejor juicio relativo a qué hacer. El problema aquí es que incluso ese mejor juicio implica un juicio.

Como vimos en el capítulo 1, la pregunta «¿Qué he de hacer?» y su respuesta, no son necesaria o particularmente éticas; las consideraciones éticas son uno de los tipos de *input* para la deliberación. Correspondientemente, la idea que acabo de exponer no tiene que ver en particular con lo ético, sino que se refiere al *input* y a la conclusión de cualquier razonamiento práctico. Una vez hemos ido más allá del razonamiento práctico y de sus conclusiones «completas», las cuestiones más profundas, que se conocen en los debate académicos como falacia naturalista o distinción *es/debe,* no pueden solventarse, ni siquiera revelarse, mediante el análisis del lenguaje.

La filosofía moral es un área de la filosofía en la que el «giro lingüístico», como se ha dado en llamarlo[6], no ha contribuido a dar una forma

---

[6] La expresión «giro lingüístico» procede del título de una compilación de artículos sobre el método filosófico editada por Richard Rorty: *The Linguistic Turn* (Chicago, University of Chicago Press, 1967) [*El giro lingüístico: dificultades metafilosóficas de la filosofía lingüística,* trad. de G. Bello Reguera (Barcelona, Paidós, 1990)].

más tratable a los problemas. Esto no niega que la filosofía moral, como otras disciplinas filosóficas, esté adecuadamente concernida por la reflexión sobre qué decimos. De hecho, a cierto nivel podría haberlo hecho mejor de lo que lo ha hecho si se hubiera ocupado más de lo que decimos. Su fallo primordial, en todos los estilos que adopta, ha sido imponer a la vida ética cierto modelo inmensamente simple, sea de los conceptos que usamos o de las reglas morales por las que hemos de guiarnos. Se habría podido poner remedio a esta deformación persistente con solo prestar atención a la gran diversidad de cosas que decimos sobre cómo nosotros y los demás vivimos nuestras vidas.

A cierto nivel, si hay que prestar atención al lenguaje, ha de ser a más aspectos suyos. Pero, a otro nivel, no es muy probable que el empeño distintivamente lingüístico llegue a buen puerto, se lleve a cabo como se lleve. Una razón para ello, como ya hemos señalado, es que no hay un conjunto claramente delimitado de expresiones lingüísticas que quepa examinar. Los teóricos han tendido a privilegiar las expresiones más generales que se usan en el debate ético: *bien, correcto, debe,* etc. Nuestro uso de estas palabras no está, por supuesto, confinado al pensamiento ético. Ese hecho no basta para condenar la investigación, pero, por al menos dos razones, esta focalización ha contribuido a ello. Una de las razones tiene que ver con el motivo para elegir esas palabras, a saber, la creencia reduccionista de que esas nociones estaban contenidas en concepciones éticas más específicas. Esto tiende a enmascarar la naturaleza real de esas concepciones y ha contribuido a ocultar una verdad que es improbable que revele una investigación puramente lingüística: una sociedad que descansa en expresiones éticas muy generales es un tipo diferente de sociedad de la que insiste en expresiones éticas más específicas. (En el siguiente capítulo me ocuparé de algunas características importantes de los conceptos éticos más específicos; en particular, del papel que desempeñan en el conocimiento ético). Una segunda razón de por qué centrarse en esos términos generales no ha aportado nada a la filosofía lingüística de la ética es que el teórico, al tratar de seleccionar sus usos relevantes, incorpora presupuestos a la investigación que no son solo ya teóricos sino también éticos. El resultado suele ser una mala filosofía del lenguaje.

En todo caso, es una mala filosofía de la ética. Las dudas sobre el *debe* y el *es* y sobre la falacia naturalista generan preocupaciones éticas genuinas y, en última instancia, metafísicas. En el núcleo de estas preocupa-

ciones se encuentra la idea de que nuestros valores no están «en el mundo», de que una descripción del mundo que no sea tendenciosa no mencionaría ningún valor, de que imponemos o proyectamos nuestros valores en nuestro entorno. Este descubrimiento, si verdaderamente lo es, puede causar desesperación, como también puede hacerlo la pérdida de un mundo teleológicamente significativo. Pero también puede verse como una liberación, y podemos hallar una forma radical de libertad en el hecho de que el mundo no pueda forzarnos a aceptar un determinado conjunto de valores en detrimento de otro.

Todas estas concepciones nos llevan a creer en una distinción, en algún tipo de distinción, entre hechos y valores. Que exista o no una distinción sólida de ese tipo constituye, sin duda, una cuestión muy seria. Nos ocuparemos de ello en el siguiente capítulo, así como de la cuestión de *qué* concepción del mundo estaría libre de valoraciones. Lo que hace falta abordar ahora es preliminar, pero importante. Muestra por qué es probable que el giro lingüístico sea de poca ayuda. Si hay una distinción fundamental entre hechos y valores, no es una proeza universal de la humanidad haberla reconocido, sino un descubrimiento, un logro de la Ilustración. Pero, en tal caso, no hay razón alguna para suponer que nuestro lenguaje ético —en la medida en que exista tal cosa de un modo bien definido— comporta *ya* esta distinción. Puede ser que no manifieste nada de ese tipo, sea sugiriendo u ocultando la distinción; de hecho, puede ser un error creer que el lenguaje incorpora creencias distintivamente metafísicas. Pero, si tiene la capacidad de aportarnos algo sobre esa cuestión, tiene que ser al menos tan probable que transmita una ilusión como que transmita una verdad, y en realidad es más probable lo primero que lo segundo, sobre todo si se tiene en cuenta lo extendida que está la ilusión en el espacio y el tiempo y lo reciente de la Ilustración. Si los valores humanos son proyección de los intereses humanos, y no una característica del «mundo», de aquí no se sigue que quepa alcanzar una descripción del mundo —o una porción de él— que esté libre de valores. (Tal vez no pueda haber tal cosa; y, si no la hay, habremos de empezar a preguntarnos qué decimos realmente al hablar de proyección. ¿Dónde se encuentra la pantalla?).

Esta puede parecer una forma un tanto paradójica de criticar el empeño lingüístico. Quienes lo han cultivado han insistido, por lo general, en la distinción entre hechos y valores utilizando la distinción entre el *debe* y el *es* para revelarla. Así, parece que o bien el lenguaje no oculta la distinción hecho/valor, o el teórico lingüístico ha conseguido

revelar el ocultamiento. Pero ninguna de esas opciones es correcta. Lo que ha sucedido es que los teóricos han introducido la distinción hecho/valor en el lenguaje en vez de descubrir cómo se revela en él. Lo que han hallado son muchas de esas nociones éticas «densas» o más específicas a las que ya me he referido, tales como la *traición,* la *promesa,* la *brutalidad* o la *valentía,* que parecen expresar una unión entre hechos y valores. La manera en que se aplican esas nociones está determinada por cómo es el mundo (por ejemplo, por cómo ha actuado alguien), y sin embargo, al mismo tiempo, su aplicación comporta normalmente una determinada valoración de la situación, de las personas o de las acciones. Además, suelen proporcionar razones para la acción (aunque no necesariamente de manera directa). Sin duda, los términos de ese tipo no toleran la distinción hecho/valor, sino que es el teórico que quiere defender esta distinción quien ha de interpretar el funcionamiento de esos términos, y lo hace tratándolos como la conjunción de dos elementos, uno fáctico y otro evaluativo, que en principio pueden separarse. El análisis más claro de ello, como suele suceder, es el de Hare: este tipo de términos implica un complejo descriptivo que lleva adosada una prescripción, que es expresión de los valores del individuo o de la sociedad. Una oración que usa uno de esos términos puede analizarse en algo así como «este acto tiene tal y tal carácter, y uno no debe realizar actos de ese carácter». Es esencial para este análisis que el carácter específico o «denso» de estos términos venga dado por el elemento descriptivo. Según este análisis, la parte valorativa la expresa el término prescriptivo polivalente *debe.*

Defenderé en el siguiente capítulo que este análisis es incorrecto, y no trataré de adelantar aquí la discusión. Me limitaré a afirmar, de nuevo, que los defensores de la distinción hecho/valor que se apoyan en medios lingüísticos introducen la distinción en el lenguaje en vez de encontrarla en él; además, esperan sin razón que la revelación localice la distinción muy próxima a la superficie del lenguaje. Pero no hay razón alguna para esperar eso. Si nos dedicamos a la tarea fraudulenta o ilusoria de leer nuestros valores en el mundo, es muy probable que nuestro lenguaje se implique profundamente.

La teoría prescriptivista pretende que la parte valorativa de esos términos densos recaiga completamente en la función prescriptiva, que puede analizarse en términos de *debe.* Sostiene que todo lo que queremos decir o pensar en el dominio de lo ético (y en las áreas evaluativas que lo trascienden) puede decirse o pensarse en función de ese término

muy general. Otros teóricos dicen cosas semejantes sobre otros términos generales: un término general, abstracto, bastará para todo. Y, dado que los términos éticos más densos son solo compuestos que incluyen este término, sobre él descansa todo. Si este tipo de análisis resulta ser erróneo y, de manera más general, si la tendencia a reducir el lenguaje ético a tales términos abstractos fracasa, entonces estos términos no bastan, y quedará margen para la idea, como ya he sugerido, de que una sociedad en la que la vida ética se entiende y conduce en tales términos generales es socialmente diferente a una en la que no es así, y las diferencias requieren una comprensión social. Si esto es así, el planteamiento lingüístico no nos ayuda en absoluto a reconocer esta posibilidad, sino que nos anima a pasarla por alto.

Resulta bastante obvio que si queremos comprender cómo funcionan los conceptos éticos, y cómo cambian, hayamos de tener algún conocimiento de las formas de organización social en las que operan. El planteamiento lingüístico no niega esto, a cierto nivel poco comprometido, pero no plantea ninguna cuestión que contribuya a alcanzar esa comprensión o, una vez adquirida, a sacarle ningún provecho filosófico. Al centrar la atención en cuestiones de análisis lógico, contribuye a ocultar este problema, como lo hace la concepción pura de la filosofía que suscribe, que enfatiza que el lenguaje es una actividad social, pero rechaza al mismo tiempo, de un modo bastante insólito, que la filosofía tenga ningún interés concreto en las sociedades[7]. Por lo menos, esta perspectiva es más apta para comprender la dimensión social e histórica del pensamiento ético que otras que lo consideran un objeto de estudio autónomo e invariable. Prestar atención a nuestro lenguaje ético, cuanto menos, puede invitarnos a que pensemos en él, y en la vida ética que expresa, como prácticas sociales que pueden cambiar. Aunque no lo haya hecho, el giro lingüístico podría ayudarnos a admitir que la comprensión ética requiere una dimensión explicativa de tipo social.

---

[7] Una excepción la encontramos en Peter Winch, *The Idea of a Social Science and Its Relations to Philosophy* (Londres, 1958; Atlantic Highlands, N. J., Humanities Press, 1970) [*Ciencia social y filosofía,* trad. de M. R. Viganó de Bonacalza (Buenos Aires, Amorrortu, 1972)], y en otros de sus escritos.

Capítulo 8

Conocimiento, ciencia y convergencia

Hasta el momento no he dicho gran cosa sobre la objetividad, aunque los primeros capítulos tenían mucho que ver con ella. Si se pudiera introducir un punto arquimédico y mostrar que la razón práctica, o los intereses humanos, implican una perspectiva ética determinada, entonces el pensamiento ético sería objetivo, en el sentido de que se le habría dado un fundamento objetivo. Estas son posibilidades —o podrían serlo— propias de la perspectiva de la razón práctica. No obstante, muy a menudo los debates en torno a la objetividad en la filosofía moral tienen un punto de partida diferente y provienen del interés en comparar las creencias éticas con el conocimiento y otros tipos de pretensiones de verdad, como las creencias científicas. Lo que aquí está en juego es otra concepción de la objetividad que es natural vincular con cuestiones como la de qué puede hacer verdaderas a las creencias éticas y la de si existe un conocimiento ético. Es en este campo de comparaciones donde cabe situar diversas distinciones entre hechos y valores.

Los debates sobre la objetividad suelen partir de consideraciones relativas al desacuerdo. Esto hace que parezca como si el desacuerdo fuera sorprendente, pero no hay ninguna razón por la que eso haya de ser así (los primeros pensadores de la tradición occidental creyeron que el conflicto era un rasgo distintivo del mundo tan evidente como la

concordia). El interés en el desacuerdo se debe, en cambio, a que ni el acuerdo ni el desacuerdo son universales. No es que el desacuerdo necesite una explicación y el acuerdo no, sino que, en diferentes contextos, el desacuerdo requiere diferentes tipos de explicación, y lo mismo sucede con el acuerdo.

El modo como comprendemos y explicamos un tipo dado de desacuerdo tiene repercusiones prácticas importantes, pues puede modificar nuestra actitud hacia los demás y la comprensión que tenemos de nuestra propia perspectiva. En las relaciones con los demás, necesitamos una idea de a qué cabe oponerse, rechazar, etc., y con qué espíritu; con respecto a nosotros mismos, el desacuerdo puede advertirnos de que podemos estar equivocados y, si aspiramos a la verdad como precisión, puede que necesitemos reformar nuestras estrategias.

No es necesario que se supere el desacuerdo. Este puede mantenerse como una característica importante y constitutiva de nuestras relaciones con los demás, y también concebirse como algo que es de esperar a la luz de las explicaciones de que disponemos de la manera en que surge ese desacuerdo. Aquí puede haber una tensión si, al mismo tiempo, sentimos que es un desacuerdo sobre asuntos muy importantes y que hay una buena explicación de por qué surge el desacuerdo. La tensión es particularmente aguda cuando el desacuerdo no es solo importante sino que también se expresa en juicios que parecen pedir el asentimiento de los demás. (Como veremos en el capítulo siguiente, el problema específico del relativismo consiste en comprender nuestro punto de vista de un modo que dé cabida a los dos extremos de esta tensión).

Entre los diferentes tipos de desacuerdo y las lecciones que podemos aprender de ellos hay una polaridad bien conocida. En un extremo está la situación de dos niños que quieren un mismo dulce o la de dos héroes que quieren a la misma esclava. El desacuerdo es práctico, y su explicación no va a arrojar muchas dudas sobre las capacidades cognitivas de las personas afectadas. Puede decirse que este tipo de casos es práctico de una forma tan primaria que es difícil que plantee un juicio sobre el que haya desacuerdo. Por supuesto, incluso en el nivel más primario existe un desacuerdo sobre *qué hacer*, pero está tan próximo al deseo y a la acción que nadie pensaría que el desacuerdo manifiesta algún fallo cognitivo o de comprensión. Es simplemente que dos personas quieren cosas incompatibles. Pero el conflicto puede que no sea tan inofensivo y, si las partes quieren resolverlo por medio de un debate coherente, en vez de regirse por la violencia, invocarán juicios más sus-

tantivos, normalmente relativos a la justicia; los niños hablarán de la equidad y los héroes de la precedencia.

Siquiera en su forma más básica, no hace falta que en estos desacuerdos ninguna de las partes piense que la otra no ha reconocido o comprendido algo, o que no dominan el lenguaje. En el polo opuesto del contraste tradicional se encuentran los desacuerdos que suscitan que una de las partes piense esto. En qué consisten estos desacuerdos depende de la teoría del conocimiento preferida por el comentador respectivo, pero suelen implicar la observación en condiciones normales de lo que el filósofo oxoniense J. L. Austin solía llamar «especímenes de tamaño moderado de bienes sólidos»*. Una característica importante de estos ejemplos es que se presupone que las partes comparten los mismos conceptos y son diestras en el reconocimiento de muebles, lápices, monedas o lo que sea.

Alrededor de estos paradigmas se han formado diversos tipos de oposiciones: entre lo práctico y lo teórico, los valores y los hechos, el *debe* y el *es*. Se ha considerado que cada una de ellas representa una diferencia fundamental en lo que significa el desacuerdo, y suele pensarse que sugieren expectativas contradictorias sobre la forma de resolverlo. Pero es un error suponer que esas oposiciones son maneras diferentes de representar una misma distinción. De hecho, es significativo que los dos ejemplos que he mencionado no se correspondan con los dos polos de cada una de estas oposiciones. La disputa sobre la atribución de un bien es ciertamente un ejemplo práctico, pero hasta que se alcanza el estadio en el que se toman en serio las exigencias de la justicia, no se trata todavía de un desacuerdo sobre el valor. Un desacuerdo sobre la percepción de un mueble es, sin duda, un desacuerdo sobre una cuestión de hecho, pero no lo es aún sobre lo que suele contrastar con lo práctico, esto es, sobre lo teórico. Clasificar esos tipos de ejemplos en una única oposición exige una elaboración mayor. De modo característico, esto se ha hecho por reducción de lo evaluativo a lo práctico y estirando lo fáctico hasta lo teórico. Ambas maniobras son de inspiración positivista, y resultan sospechosas. No es sorprendente que algunos filósofos duden hoy en día de que se pueda construir una distinción fundamental siguiendo el patrón tradicional[1].

---

* Exactamente, «*middle-sized dry goods*». *[N. del T.]*.

[1] La ruta más conocida hacia la reducción de lo evaluativo a lo práctico pasa por la noción de lo prescriptivo. Esta estrategia fue criticada en el capítulo anterior.

Acepto que aquí no hay una única distinción en juego. También acepto que las formulaciones más positivistas que han definido cada polo de tal distinción son erróneas. Pero creo que en ética hay que encontrar una diferencia genuina y profunda, y también —y esta es otra idea— que la diferencia basta para motivar alguna versión de la sensación (recurrente en sí misma, si no tradicional) de que la ciencia tiene buenas opciones de ser más o menos lo que parece, una exposición teórica sistemática de cómo es el mundo, mientras que el pensamiento ético no tiene ninguna opción de ser todo lo que parece. La tradición tiene razón, además, no solo en pensar que existe esa distinción, sino también en que podemos comprenderla comprendiendo los desacuerdos. Sin embargo, no es cuestión del grado de desacuerdo existente, ni siquiera de qué métodos permiten resolver los desacuerdos, si bien no cabe duda de que esto proporciona muchas consideraciones relevantes. La diferencia fundamental reside, sobre todo, en nuestra comprensión reflexiva de las mejores expectativas que podamos concebir coherentemente para eliminar el desacuerdo en ambas áreas. Es cuestión de cuál sería, en las condiciones más favorables, la mejor explicación del fin del desacuerdo; la explicación —como diré de ahora en adelante— de la convergencia.

La diferencia fundamental es entre lo ético y lo científico. Espero explicar por qué un extremo de la oposición ha de denominarse «lo científico» en vez de, pongamos por caso, «lo fáctico». El otro extremo se denomina «lo ético» porque lo ético es nuestro objetivo, y exigiría un largo debate ampliar el campo o estrecharlo. No lo llamamos «lo evaluativo» porque esta etiqueta también cubre, por lo menos, el área del juicio estético, el cual plantea muchas cuestiones de por sí. Ni lo llamamos «lo normativo» porque este es un término que cubre solo parte del interés de lo ético (a grandes rasgos, la parte concerniente a las reglas) y es natural que se extienda también a cosas tales como el derecho, que plantea cuestiones diferentes. Lo que resulta aún más significativo: no lo llamamos tampoco «lo práctico», pues esto desplazaría en gran medida el problema, por una razón que ya hemos señalado en el examen de las prescripciones y de la distinción *es/debe*. No es difícil conceder que existe una distinción entre lo práctico y lo no práctico. No se puede negar la existencia del razonamiento o de la deliberación prácticos, los cuales no tienen nada que ver con pensar cómo son las cosas. *Obviamente* son cosas diferentes, y esta es la razón por la que el positivismo había validado la distinción tradicional por medio de la reducción de lo

evaluativo a lo práctico. Pero esta reducción es errónea y enmascara la dificultad del problema[2].

La idea fundamental que está en el trasfondo de la distinción entre lo científico y lo ético, expresada en términos de convergencia, es muy simple. En la investigación científica, idealmente, ha de haber convergencia en una respuesta, donde la mejor explicación de la convergencia implica la idea de que la respuesta representa cómo son las cosas; en el área de lo ético, al menos en términos generales, no hay tal expectativa de convergencia. La distinción no pivota sobre una diferencia relativa a si la convergencia realmente se dará o no, y es importante que la discusión no dependa de esto. Podría resultar que hay convergencia en las perspectivas éticas, por lo menos entre los seres humanos. El sentido de contrastar lo ético y lo científico es que, incluso si se da la convergencia en lo ético, será erróneo pensar que se debe a que la convergencia responde a cómo son las cosas en realidad, mientras que la convergencia en las ciencias sí que podría explicarse así. Entre otras cosas, esto significa que nuestra comprensión de la convergencia, o de la falta de ella, es diferente en cada uno de los dos casos.

Volveré a la cuestión de cómo podemos entender la convergencia en ética. Pero, en primer lugar, tenemos que afrontar determinados argumentos que sugieren que, expresada en esos términos, la distinción no se sostiene. Esta distinción puede tener dos orígenes diferentes. En una de sus versiones, que considera que la convergencia procede de cómo son las cosas en realidad, es vista como una noción vacía. Según otra versión, esta noción de convergencia no es vacía, pero sirve tanto para los casos éticos como para los científicos (lo que significa que la noción tiene un contenido, pero no aporta nada a la distinción).

Ya he dicho que la razón de ser de la distinción y de su explicación en términos de convergencia no depende de la cuestión de si la convergencia se produce o no. En el campo científico, sin embargo, sería irreal desconectar por completo esas ideas de la cuestión de cómo tiene que entenderse la historia de la ciencia occidental desde el siglo XVI. La concepción del progreso científico en términos de convergencia no

[2] Véase Wiggins, «Truth, Invention, and the Meaning of Life» y «Deliberation and Practical Reason».

puede separarse de la historia de la ciencia occidental porque es esta historia la que ha hecho más por fomentarla. Es muy difícil negar que la historia de la ciencia occidental muestra un grado considerable de convergencia; aunque se ha pretendido que esta apariencia no tiene ninguna significación real porque se trata de un artefacto cultural, un producto de cómo hemos elegido narrar la historia de la ciencia. Richard Rorty ha escrito lo siguiente:

> Resulta menos paradójico [...] que la física apele a la noción clásica de «describir mejor lo que ya estaba ahí». Esto no se debe a profundas consideraciones epistemológicas o metafísicas, sino solo a que, cuando nos contamos las típicas historias de progresistas sobre cómo nuestros antepasados han escalado paso a paso la montaña sobre cuya cima (posiblemente falsa) nos encontramos, hace falta que mantengamos constantes ciertas cosas a lo largo del relato [...]. La física es el paradigma del «descubrimiento» solamente porque es difícil (al menos en Occidente) contar una historia de los universos cambiantes sobre el trasfondo de una ley moral o un canon poético invariables, mientras que es muy fácil contar la historia contraria[3].

Hay dos fallos notables en esta descripción del éxito científico y de lo que este significa. El primero es su actitud hacia el hecho de que es fácil contar un tipo de historia y difícil contar el otro. *¿Por qué* es tan convincente la imagen del mundo «que ya estaba ahí», que nos ayuda a controlar las descripciones que damos de él? La posición de Rorty tiene que dar algún tipo de explicación de esto, pero no la proporciona. Si la referencia a «Occidente» supone una explicación cultural o antropológica, no se acaba de ver en qué consistiría; en realidad, no se acaba de ver cuál podría ser si no se presupone un mundo físico preexistente en el que los seres humanos aparecen y desarrollan sus culturas.

La razón de que un presupuesto de este tipo subyazca a cualquier explicación de lo que hacemos conduce directamente al segundo fallo de

---

[3] Richard Rorty, *Philosophy and the Mirror of Nature* (Princeton, Princeton University Press, 1980), págs. 344-345 [*La filosofía y el espejo de la naturaleza*, trad. de J. Fernández Zulaica (Madrid, Cátedra, 1989); pero el fragmento citado es traducción propia]. He examinado con cierto detalle las ideas de Rorty en una reseña de su libro *Consequences of Pragmatism* (Minneapolis, University of Minnesota Press, 1982) [*Consecuencias del pragmatismo*, trad. de J. M. Esteban Cloquell (Madrid, Tecnos, 1996)], publicada en *New York Review of Books*, 28 de abril de 1983.

la posición de Rorty: se refuta a sí misma. Si la historia que cuenta fuera cierta, no habría ninguna perspectiva desde la que pudiera exponerla de esta forma. Si es sumamente conveniente decir que la ciencia describe lo que ya está ahí, y si no están en juego asuntos metafísicos o epistemológicos profundos, sino que es solo una cuestión de conveniencia (es «solamente por» esto por lo que hablamos como lo hacemos), entonces lo que todo el mundo habría de decir, incluido Rorty, es que la ciencia describe un mundo que ya estaba ahí. Pero Rorty nos recomienda no decir eso, y simultáneamente, al insistir, en *contraposición* a eso, en que hablemos de lo que resulta más conveniente decir, está tratando de volver a ocupar una posición trascendental externa a la actividad y el lenguaje humanos, que es precisamente a lo que quiere que renunciemos[4].

Otra objeción, más eficaz, se encuentra en la tesis negativa que defienden Rorty y otros autores, según la cual ninguna convergencia en la ciencia, pasada o futura, se podrá explicar de una manera adecuada por referencia a cómo es el mundo a causa de una dificultad insoluble en la noción de «el mundo» como algo que pueda fijar la creencia. Se plantea un dilema. Por un lado, «el mundo» puede caracterizarse en términos de nuestras creencias presentes sobre lo que contiene: es un mundo de estrellas, personas, hierba o mesas. Cuando concebimos «el mundo» de esta manera podemos decir que nuestras creencias sobre el mundo se ven afectadas por el mundo, en el sentido de que, por ejemplo, nuestras creencias sobre la hierba se ven afectadas por la hierba, pero no hay nada iluminador o sustantivo en esto; nuestra concepción del mundo como el objeto de nuestras creencias no puede hacer más que repetir las creencias que consideramos que lo representan. Si, por otro lado, tratamos de formarnos alguna idea de un mundo que es previo a cualquier descripción de él —el mundo que todos los sistemas de creencia y de representación tratan de representar—, entonces obtenemos una noción vacía de algo totalmente inespecífico e inespecificable[5].

---

[4] A menudo se confunde el pragmatismo que podría llamarse empírico y el trascendental. Problemas semejantes se plantean con respecto al segundo Wittgentein; véase mi artículo «Wittgenstein y el idealismo», recogido en *La fortuna moral;* y Jonathan Lear, «Leaving the World Alone», *Journal of Philosophy,* 79 (1982).

[5] Rorty, «El mundo felizmente perdido», en *Consecuencias del pragmatismo,* pág. 14 [ed. inglesa]. Véase también Donald Davidson, «The Very Idea of a Conceptual Scheme», en *Proceedings and Addresses of the American Philosophical Assotiation,* 67 (1973-1974) [«De la idea misma de un esquema conceptual», en *De la verdad y la interpretación,* trad. de G. Filippi (Barcelona, Gedisa, 1990), págs. 189-203].

Así, en cualquier caso, no llegamos a una noción de «el mundo» que realice la función esperada.

Cada cuerno del dilema reúne todas nuestras representaciones del mundo, incluyéndolas todas en un caso y excluyéndolas todas en el otro. Pero hay una tercera posibilidad más útil: la de que nos formemos una concepción del mundo que estaba «ya ahí» en términos de algunas de nuestras creencias y teorías, pero de no todas. Al reflexionar sobre el mundo que está ahí *en todo caso,* independientemente de nuestra experiencia, no tenemos que centrarnos en primer lugar en el objeto de nuestras creencias, sino en la forma en que estas representan su objeto. Podemos seleccionar, de entre las creencias y características de nuestra imagen del mundo, aquellas que podemos afirmar razonablemente que representan el mundo de un modo que posea el máximo grado de independencia con respecto a nuestra perspectiva y sus peculiaridades. Si podemos llevar a cabo esta tarea, la imagen de las cosas resultante puede denominarse la «concepción absoluta» del mundo[6]. A partir de esta concepción, podemos tener la esperanza de explicar cómo es posible alcanzar esa misma concepción, y también cómo son posibles otras representaciones perspectivistas.

Esta noción de una concepción absoluta puede contribuir a hacer efectiva una distinción entre «el mundo en tanto que independiente de nuestra experiencia» y «el mundo tal y como se nos aparece». Lo hace concibiendo «el mundo tal y como se nos aparece» como «el mundo tal y como se nos aparece a nosotros en particular»; la concepción absoluta será, correspondientemente, una concepción del mundo accesible a cualquier investigador, incluso si fuera muy distinto a nosotros. Lo que cuenta como una diferencia relevante entre nosotros, y lo que contará como «para nosotros» en muchos niveles de descripción, se explicará sobre la base de la concepción misma; por ejemplo, podremos explicar por qué un tipo de observador puede hacer observaciones que son imposibles para uno de otro tipo. Es de una importancia fundamental que esas ideas se refieran a la ciencia, y no a todos los tipos de conocimiento. Podemos *saber* cosas cuyo contenido es perspectivista: podemos saber que la hierba es verde, por ejemplo, aunque con toda seguridad *verde,* y

---

[6] Esto ya se ha mencionado, capítulo 6, nota 14. Véase también N. Jardine, «The Possibility of Absolutism», en D. H. Mellor (ed.), *Science, Belief, and Behaviour: Essays in Honour of R. B. Braithwaite* (Nueva York, Cambridge University Press, 1980); y Colin McGinn, *The Subjective View* (Oxford, Clarendon Press, 1983).

probablemente *hierba,* no sean conceptos accesibles a todo observador competente del mundo y no figurarán en la concepción absoluta. (Como veremos, las personas pueden saber cosas según una perspectiva aún más local que esa). La cuestión no es dar cuenta del conocimiento; el contraste con el valor no ha de expresarse en términos de conocimiento, sino en términos de ciencia. Nuestro objetivo es esbozar la posibilidad de una convergencia característica de la ciencia, de la que pueda decirse con razón que es una convergencia concerniente a cómo son las cosas (en todo caso).

Esa posibilidad, como ya he explicado, depende fundamentalmente de las nociones de explicación. La sustancia de la concepción absoluta (en oposición a esas ideas vacuas o tendentes a desvanecerse de «el mundo» que se presentaron antes) se encuentra en la idea de que se podría explicar de modo no vacuo cómo son posibles ella misma y las diversas concepciones perspectivistas del mundo. Uno de los rasgos más importantes de la ciencia moderna es su contribución a la explicación de cómo criaturas con nuestros orígenes y características pueden comprender un mundo dotado de las propiedades que esta misma ciencia le adscribe. En este sentido, son esenciales los logros de la biología evolutiva y de las neurociencias, y no son vacuas sus nociones de explicación. Es verdad, no obstante, que esas explicaciones no pueden operar por completo al nivel de la concepción absoluta, porque tienen que explicar fenómenos psicológicos y sociales, tales como las creencias, las teorías y las concepciones del mundo, y puede ser poco razonable suponer que, a su vez, podrían caracterizarse adecuadamente en términos no perspectivistas. La cuestión de en qué medida esto sería posible es una cuestión filosófica fundamental. Pero, incluso si se admite que las explicaciones de estos fenómenos tienen que conservar algún grado de perspectivismo, esto no significa que la noción de una concepción absoluta no sea operativa. Será una concepción consistente en un compuesto de elementos no perspectivistas accesible a cualquier investigador competente, independientemente de su constitución, y también contribuirá a explicarnos a nosotros, aunque no necesariamente a esos investigadores foráneos, cosas como nuestra capacidad de comprender esa concepción. Puede que nos aporte más, pero esto es ya suficiente para dar sustancia a la idea de «el mundo» y para refutar el primer tipo de objeción a la distinción entre lo científico y lo ético en términos de convergencia posible.

La objeción opuesta sostiene que la idea de «converger con respecto a cómo son las cosas», en cierto grado, es también accesible a la ética. Esto se puede apreciar sobre todo en los conceptos éticos sustantivos o densos que he mencionado a menudo. Pueden aducirse muchos ejemplos exóticos de este tipo procedentes de otras culturas, pero bastará con quedarnos en la nuestra: *cobarde, mentira, brutalidad, gratitud,* etc. Es característica su vinculación con razones para actuar. La aplicación de un concepto de este tipo suele proporcionar a alguien una razón para actuar, aunque esta razón no tiene por qué ser decisiva y puede ser contrarrestada por otras razones, como vimos en relación con su función en el razonamiento práctico en el capítulo 1. Por supuesto, qué razón exacta para realizar la acción se ofrezca, y a quién, depende de la situación, lo cual puede estar regido por un concepto ético u otro, pero su conexión general con la acción resulta bastante clara. En suma, podemos decir que estos conceptos «guían la acción».

Al mismo tiempo, el mundo guía su aplicación. Un concepto de este tipo puede aplicarse bien o mal, y las personas que lo han adquirido pueden estar de acuerdo en si se aplica o no a una situación nueva. En muchos casos, el acuerdo será espontáneo, aunque en otros casos tendrá cabida el juicio y la comparación. Ciertos desacuerdos marginales pueden resultar irresolubles, pero esto no significa que el uso del concepto no esté regido por los hechos o por la percepción del mundo que tiene quien lo aplica (como sucede con otros conceptos que no son completamente precisos, los desacuerdos marginales pueden mostrar cómo su uso *está* controlado por los hechos). Podemos decir, pues, que la aplicación de esos conceptos está guiada por el mundo y, al mismo tiempo, guía la acción. Pero, ¿cómo pueden darse ambas cosas a la vez?

La teoría prescriptivista examinada en el capítulo anterior proporciona una respuesta muy simple a esta pregunta. Según esa teoría, cualquier concepto de este tipo puede analizarse en elementos descriptivos y prescriptivos: está guiado por el mundo en cuanto al contenido descriptivo, pero incorpora una señal prescriptiva. Si a causa del primer rasgo es guiado por el mundo, el segundo hace que guíe la acción. Algunas de las dificultades de esta imagen tienen que ver con el elemento prescriptivo y con cómo se supone que este guía la acción en el sentido relevante (ordenarme a mí mismo hacer algo no es un modelo obvio para que reconozca que tengo una razón para hacerlo). Pero la objeción más significativa se aplica a la otra mitad del análisis. El prescriptivismo pretende que lo que gobierna la aplicación del concepto al mundo es el

elemento descriptivo y que el interés evaluativo del concepto no desempeña ningún papel. Todo el *input* para su uso es descriptivo, de igual manera que todo el aspecto evaluativo es *output*. Se sigue que, para cualquier concepto de este tipo, podríamos producir otro que seleccionase exactamente los mismos rasgos del mundo, pero que funcionara simplemente como un concepto descriptivo, carente de cualquier fuerza prescriptiva o evaluativa.

Contra esto, los críticos han defendido eficazmente que no hay ninguna razón para creer que vayamos a poder encontrar un equivalente descriptivo[7]. La forma en la que «pasamos» de una a otra aplicación de un concepto se da en función del tipo de interés que el concepto represente, y no deberíamos presuponer que nos es posible detectar la manera en que las personas hacen esta transición sin compartir la perspectiva evaluativa inherente a este tipo de concepto. Un observador perspicaz puede llegar a comprender y anticipar el uso del concepto sin compartir los valores de las personas que lo usan; esta es una cuestión

---

[7] En especial, John McDowell, «Are Moral Requirements Hypothetical Imperatives?», *Proceedings of the Aristotelian Society*, vol. supl. 52 (1978); y «Virtue and Reason», *Monist*, 62 (1979). McDowell está interesado, sobre todo, en los estados mentales y las motivaciones de una persona virtuosa, pero entiendo que su posición tiene implicaciones más generales que discuto en este texto. La idea de que podría ser imposible seleccionar un concepto evaluativo a menos que alguien comparta su interés evaluativo es una idea esencialmente wittgensteiniana. La primera vez que escuché esto fue de boca de Philippa Foot e Irish Murdoch en un seminario en los años cincuenta. Para la aplicación a la ética de ideas procedentes de la filosofía del segundo Wittgenstein, véase, por ejemplo, Hanna F. Pitkin, *Wittgenstein and Justice* (Berkeley, University of California Press, 1972) [*Wittgenstein: el lenguaje, la política y la justicia*, trad. de R. Montoro Romero (Madrid, CEPC, 1984)], y Sabina Lovibond, *Realism and Imagination in Ethics* (Minneapolis, University of Minnesota Press, 1983). Para una reflexión más general que debe mucho a Wittgenstein, véase Stanley Cavell, *The Claim of Reason* (Nueva York, Oxford University Press, 1979) [*Reivindicaciones de la razón*, trad. de D. Ribes Nicolás (Madrid, Síntesis, 2003)], esp. partes 3 y 4. El punto de vista personal de Wittgenstein respecto a las cuestiones éticas es otro tema; véase su artículo póstumo, «A Lecture on Ethics», *Philosophical Review*, 74 (1965) [*Conferencia sobre ética*, trad. de Fina Birulés (Barcelona, Paidós, 1995)]; Rush Rhees, «Some Developments of Wittgenstein's View of Ethics», *Philosophical Review*, 74 (1965); B. F. McGuinnes, «The Mysticism of the Tractatus», *Philosophical Review*, 75 (1966). El propio McDowell extrae importantes consecuencias para la filosofía de la mente, al rechazar el modelo «creencia y deseo» de la acción racional. Yo no suscribo estas consecuencias, pero no trataré de discutirlas aquí. Algunas de las consideraciones que aparecen hacia el final de este capítulo, sobre las diferencias entre las creencias éticas y las percepciones sensoriales, están estrechamente vinculadas.

importante, a la que volveré más adelante. Pero, al anticipar imaginativamente el uso del concepto, el observador también capta imaginativamente su aspecto evaluativo. No puede permanecer al margen de los intereses evaluativos de la comunidad que está observando y contentarse con seleccionar el concepto solo como un instrumento que le permite dividir, de un modo un tanto extraño, determinados rasgos neutrales del mundo.

Es muy razonable, y ciertamente plausible, que haya conceptos éticos que planteen estas exigencias a la comprensión. No hace falta que esto sea mucho más que posible para que desempeñe un papel importante en ese debate, al recordar a la filosofía moral cómo pueden ser las exigencias de una filosofía del lenguaje adecuada o de una filosofía de la explicación social. Si no es solo posible, sino también razonable, es aconsejable que la filosofía moral considere qué tiene que decirse si esto es verdad.

El observador empático puede seguir la práctica de las personas que observa; puede relatar, anticipar e incluso tomar parte en debates sobre el uso que estas personas hacen de su concepto. Pero, como con otros conceptos de quienes observa relativos a la religión, por ejemplo, o a la brujería, puede no identificarse completamente con el uso del concepto: puede que ese concepto no sea realmente como el suyo propio[8]. La posibilidad de un observador perspicaz que no se identifica por completo plantea la importante cuestión de si podemos decir que quienes aplican adecuadamente este tipo de conceptos éticos poseen un conocimiento ético.

Supongamos, de manera artificial, que estamos tratando con una sociedad homogénea en el grado máximo y con una tendencia mínima a la reflexión general; sus miembros, sin excepción, sencillamente usan determinados conceptos éticos de este tipo. (Podemos denominarla la sociedad «hipertradicional»). ¿Qué implicaría que tengan conocimiento ético? Según las mejores teorías sobre el conocimiento proposicio-

---

[8] McDowell («Virtue and Reason») acepta esta posibilidad, pero no extrae ninguna consecuencia al respecto e ignora por completo el conflicto intercultural. Por un lado, conecta el escepticismo relativo a la objetividad en ética con lo que, de manera reveladora, denomina «cientifismo filisteo» y, por otro, con una patología filosófica —el vértigo ante las prácticas sin fundamento—. Dejando de lado su actitud hacia las ciencias, McDowell parece que se interesa poco incluso por la historia y no dice nada sobre las diferencias de puntos de vista que surgen a lo largo del tiempo. Es significativo que, en un examen de las virtudes en el que se refiere principalmente a Aristóteles, tome como ejemplo la amabilidad, que no es una virtud aristotélica.

nal[9], tendrían que creer los juicios que hacen; esos juicios habrían de ser verdaderos, y sus juicios tendrían que satisfacer una condición suplementaria, que se ha debatido mucho en la filosofía del conocimiento y que puede resumirse diciendo que las dos primeras condiciones no tienen que estar conectadas de manera accidental: dado el modo en el que las personas han realizado sus indagaciones, no tiene que ser accidental que la creencia adquirida sea verdadera y, si la verdad sobre el asunto fuera otra, entonces habrían adquirido una creencia diferente, que sería verdadera en esas otras circunstancias. Así, puedo saber, mirándolo, que al tirar el dado ha salido un 6, y esto implica, en líneas generales[10], la tesis de que si hubiera salido un 4, yo me habría formado la creencia, al mirarlo, de que había salido un 4 (las situaciones alternativas que cabe considerar se han de restringir a las que se parecen más o menos la situación real). En palabras de Robert Nozick, podemos decir que el tercer requisito —que comporta mucha más elaboración de la que he sugerido— es que la creencia «rastree la verdad».

Los miembros de la sociedad hipertradicional aplican sus conceptos densos y, al hacerlo, formulan diversos juicios. Si puede decirse con razón que cualquiera de esos juicios es verdadero, entonces sus creencias pueden rastrear la verdad, dado que, si las circunstancias resultan no ser como parecía, pueden retirar sus juicios y formular otros alternativos más apropiados, etc. Todos dominan esos conceptos y pueden percibir los acaecimientos personales y sociales a los que se aplican. Si hay una verdad aquí, sus creencias pueden rastrearla. La cuestión que resta por abordar es la de si estos juicios pueden ser verdaderos.

Puede plantearse una objeción al hecho de decir que lo son. Si son verdaderos, el observador puede decir de modo correcto que lo son; si F representa uno de estos conceptos, puede decir que «La afirmación del jefe "El chico es F" es verdadera». Luego ha de poder decir, por sí mismo, «el chico es F». Pero no está preparado para ello, porque F no es uno de sus conceptos.

---

[9] El análisis más sutil e ingenioso del conocimiento proposicional que conozco es el de Robert Nozick, en el capítulo 3 de *Explicaciones filosóficas*. Algunas características fundamentales de la posición de Nozick, en especial el uso de los condicionales subjetivos, fueron anticipadas por Fred Dretske, como el propio Nozick reconoce (nota 53, pág. 630, donde se citan las referencias [ed. inglesa]).

[10] ¿Cómo de generales? Tal vez yo no pueda leer cuatro puntos como un 4, aunque pueda leer seis puntos como un 6. ¿Qué pasa si solo puedo leer seis puntos como un 6 y leo todo lo demás como un no 6?

¿Es sólida esta objeción? Descansa en el siguiente principio: A no puede decir con razón que B dice la verdad al emitir S, a menos que A pueda también decir algo equivalente a S. Esto puede parecer que se sigue de un principio básico de la teoría de la verdad, el *principio de desentrecomillado*[11], según el cual *P* es verdadero si y solo si *P*. Pero la aplicación de ese principio no es tan simple cuando se decide sobre las afirmaciones de otras personas. Por poner un ejemplo ingenuo: podemos imaginar determinada jerga escolar que usa nombres especiales para diversos objetos, lugares e instituciones de la escuela. La regla es que esas palabras solo las usan apropiadamente los miembros de la escuela, y esta regla la acepta y comprende un grupo externo a la escuela (tendría que ser así, si realmente existiera *esa* regla). Las personas saben que si usan esos términos se les tomará por miembros de esa escuela o bien se les criticará, etc. Supongamos que en esta jerga «Chabolo» era el nombre de algún edificio de la escuela. Según las reglas que estamos imaginando, un observador (en tanto que él mismo y no interpretando un papel) no podría decir con propiedad que «Robertson está en el Chabolo». Pero podría decir que «Smith dijo que "Robertson está en el Chabolo"» y podría añadir a continuación «y lo que Smith dijo es verdad». (Aunque no es necesario para el argumento, en realidad parece bastante natural que vaya un poco más lejos y diga «Smith afirmó, y es verdad, que Roberts estaba en el Chabolo»).

En este caso simple, es verdad que el observador dispone de otros términos que refieren a las mismas cosas que los términos de la jerga. Presumiblemente, así lo hacen también los usuarios locales; pero hay otros ejemplos en los que esto no es así, como en el caso de las lenguas en las que los hombres y las mujeres usan nombres diferentes para la misma cosa. En el caso de la escuela, tanto el observador como los

---

[11] Alfred Tarski, «The Concept of Truth in Formalized Languages», en *Logic, Semantics, Meta-Mathematics* (Indianápolis, Hackett Publishing Co., 1981). Sobre el tema que nos ocupa, véase David Wiggins, «What Would be a Substantial Theory of Truth?», en Zak van Straaten (ed.), *Philosophical Subjects: Essays Presented to P. F. Strawson* (Nueva York, Oxford University Press, 1980). El análisis de Wiggins plantea otra cuestión, la de si el observador podría incluso comprender qué significan las oraciones, a menos que pueda aplicarles una fórmula de la verdad como desentrecomillado. En esto, está influido por Donald Davidson, «Truth and Meaning», *Synthese*, 17 (1967) [«Verdad y significado», en L. M. Valdés Villanueva (ed.), *La búsqueda del significado* (Madrid, Tecnos, 1991)]. El hecho de que pueda haber un observador empático pero no identificado muestra que no es imposible comprender algo sin estar dispuesto a aseverarlo.

usuarios locales tienen medios verbales para distinguir lo que hace que una determinada afirmación sea verdadera en la jerga de lo que, en contraste, la hace apropiada para una persona particular. Que el género del hablante determine el término que hay que usar es más complicado. En el caso de los conceptos éticos densos, es aún más difícil, porque el observador no tiene un término que seleccione exactamente las mismas cosas que selecciona el término de los usuarios locales y que, al mismo tiempo, sea completamente independiente del interés que modela el uso local. (Por supuesto, el observador tiene una expresión como «lo que llaman F», y es importante que pueda usarla, aunque no sea independiente del término local: el manejo inteligente que haga de ella muestra que puede comprender el uso local del término, aunque no pueda usarlo él mismo).

No obstante, con independencia de las diferencias con el ejemplo simple de la jerga escolar, podemos considerar el caso del concepto ético como un ejemplo un poco más profundo de lo mismo. En ambos casos, hay una condición que tiene que satisfacerse si se quiere hablar de cierta manera, una condición que satisfacen los usuarios locales pero no el observador, y en ambos casos es cuestión de pertenecer a determinada cultura. Cuando comparamos esos casos entre sí, y ambos con la situación en la que el vocabulario se ve afectado por el género del hablante, podemos comprender por qué el observador no puede decir lo que dicen los usuarios locales, y podemos ver igualmente que nada le impide reconocer que lo que dicen puede ser verdadero. Así, el principio de desentrecomillado no nos conduce a la conclusión de que las afirmaciones de los usuarios locales, que incorporan sus conceptos éticos densos, no puedan ser verdaderas.

Existe otro argumento a favor de la conclusión de que las afirmaciones de los usuarios locales pueden no ser verdaderas. Este argumento defiende tajantemente que pueden ser falsas: no porque los usuarios locales puedan estar equivocados de maneras que ellos mismos podrían reconocer, sino porque, visto desde fuera, parece que un fragmento entero del discurso local es erróneo. Los teóricos de diferentes ámbitos han debatido mucho esta posibilidad. Los antropólogos sociales se han preguntado si, en nuestros términos, hemos de considerar erróneas las concepciones rituales o mágicas o, por el contrario, considerarlas concepciones que operan a un nivel diferente, no conmensurable con nues-

tras ideas científicas. Con independencia de lo que pueda decirse de modo más general, es difícil negar que la concepción mágica es, por lo menos, una concepción causal, con implicaciones que se solapan con las concepciones científicas de la causalidad[12]. En la medida en que esto es así, las concepciones mágicas pueden considerarse falsas vistas desde fuera, y así nadie tomaría por verdadera ninguna afirmación que defienda la influencia de la magia, incluso si puede haber utilizado correctamente todos los criterios locales que le permiten invocar un elemento concreto de influencia mágica. Los criterios locales no llegan a todo lo que entrañan esas invocaciones. En este tipo de casos, el problema de conceder que las afirmaciones locales son verdaderas es el opuesto al que se ha considerado antes. La tesis precedente pretendía que las nociones de los usuarios locales eran tan diferentes a las del observador que este no podría aseverar lo que ellos aseveraban. Ahora el problema es que sus afirmaciones pueden implicar nociones lo bastante semejantes a algunas de las del observador que este no podría negar lo que aseveran.

Podemos advertir así que las afirmaciones éticas locales plantean esta dificultad. Según esta interpretación, las afirmaciones de los usuarios locales implican algo que el observador puede formular en sus términos y que rechaza: que es *correcto,* o *está bien,* hacer cosas que él cree que no sea correcto, o que esté bien, hacer. El prescriptivismo concibe las cosas de esta forma. Las afirmaciones locales implican, junto con su contenido descriptivo, un *debe* polivalente. Hemos rechazado la parte descriptiva de este análisis: ¿hay alguna razón para aceptar la parte prescriptiva?

Por supuesto, hay un sentido mínimo en el que los locales creen que «está bien» actuar como ellos actúan, lo que no está solo implícito, sino también manifiesto, en su manera de vivir. Decir que «creen que está bien» a este nivel no significa mencionar ningún otro juicio suyo debatible; es solo dejar constancia de su práctica. ¿Tenemos que estar de acuerdo en que hay un juicio, que se expresa mediante el uso de alguna noción moral universal, que ellos aceptan y que el observador puede rechazar sin ambages?

No creo que sea necesario aceptar esta idea. Para ser más exactos: no creo que podamos decidir si la aceptamos o no hasta que nos repre-

---

[12] Véase John Skorupski, *Symbol and Theory* (Nueva York, Cambridge University Press, 1976).

sentemos de una manera más general toda la cuestión; este no es un debate que en sí mismo pueda hacernos llegar a conclusiones más generales. La cuestión fundamental es la de cómo tenemos que concebir las relaciones entre la práctica y la reflexión. El tipo muy general de juicio que está aquí en cuestión —un juicio que usa un concepto muy general— es, en esencia, producto de la reflexión, y resulta problemático cuando alguien se distancia de las prácticas de la sociedad y del uso que hace de esos conceptos y se pregunta si está bien ir por este camino, si es así como hay que evaluar las acciones y si es correcto admirar los tipos de carácter que admira. En muchas sociedades tradicionales hay algún grado de cuestionamiento reflexivo y de crítica, y esto es algo importante. Es por mor del argumento, por distinguir las cuestiones, que he usado la idea de una sociedad hipertradicional en la que la reflexión está completamente ausente.

En relación con esta sociedad, la pregunta que se plantea ahora es: ¿La práctica de la sociedad, en particular, los juicios emitidos por sus miembros, implican respuestas a cuestiones reflexivas sobre esa práctica, cuestiones que ellos nunca se han plateado? Algunos juicios emitidos por miembros de una sociedad sí tienen implicaciones a un nivel más general o teórico que ellos nunca han considerado. Esto será verdadero de sus juicios mágicos si se entienden como afirmaciones causales; es verdad de sus juicios matemáticos y de sus juicios sobre las estrellas. Tenemos cierta libertad para interpretar lo que dicen como la expresión de juicios matemáticos u opiniones sobre las estrellas; pero, si consideramos efectivamente que emiten esos juicios y expresan esas opiniones, sus afirmaciones tendrán implicaciones más generales. Si una afirmación expresa una opinión sobre las estrellas, se sigue que puede ser contradicha por otra opinión sobre las estrellas.

Hay dos modos diferentes en los que podemos ver las actividades de la sociedad hipertradicional y que dependen de diferentes modelos de práctica ética. Uno de ellos puede denominarse modelo «objetivista». Según este modelo, pensaremos que los miembros de la sociedad tratan, a su manera local, de descubrir la verdad acerca de los valores, una actividad en la que nosotros y los demás seres humanos, y también quizá criaturas que no son seres humanos, estamos todos comprometidos. Advertiremos, así, que sus juicios tienen implicaciones generales de manera semejante a como entendemos que las afirmaciones primitivas sobre las estrellas tienen implicaciones susceptibles de ser contradichas por afirmaciones más sofisticadas sobre las estrellas. De acuerdo

con el otro modelo, sin embargo, concebiremos sus juicios como parte de su modo de vida, un artefacto cultural en el que habitan (aunque no lo hayan construido conscientemente). Según este modelo, no objetivista, adoptaremos una posición diferente con respecto a las relaciones entre esa práctica y la reflexión crítica. No estaremos dispuestos a pensar que el plano de la reflexión está presente de forma implícita y no diremos que sus juicios, tal y como son planteados, tienen esas implicaciones.

La elección entre estos dos modos diferentes de entender sus actividades determinará que afirmemos que las personas de la sociedad hipertradicional tienen conocimiento ético o no. Es importante tener muy claro de qué tipo de conocimiento ético se trata. Es un conocimiento que está implicado en su elaboración de juicios en los que usan sus conceptos densos. No nos preguntamos si muestran conocimiento *al usar esos conceptos en vez de otros:* este sería un asunto al nivel reflexivo. En este sentido, la pregunta de «¿Posee la sociedad un conocimiento ético?» resulta muy ambigua. La referencia colectiva a la sociedad nos incita a adoptar la perspectiva según la cual sus representaciones éticas se comparan con las de otras sociedades, y este es el nivel reflexivo, del que no hay duda de que sus miembros no poseen conocimiento. La pregunta tiene otro sentido, en el que plantea si los miembros de la sociedad podrían, al ejercitar sus conceptos, expresar un conocimiento sobre el mundo al que los aplican, y la respuesta en este caso podría ser afirmativa.

El resultado interesante de este debate es que la respuesta será afirmativa si adoptamos un punto de vista no objetivista de las actividades éticas de los miembros de la sociedad: muchos de ellos tendrán conocimiento si utilizan sus conceptos con discernimiento, usan los criterios apropiados, etc. Pero, según el punto de vista objetivista, no tienen conocimiento, o al menos es muy improbable que lo tengan, dado que sus juicios tienen implicaciones más amplias que ellos nunca han considerado de modo reflexivo, y tenemos toda la razón en creer que, si se tienen en cuenta, el uso tradicional de los conceptos éticos resultará muy afectado.

La perspectiva objetivista, aunque niega el conocimiento a la sociedad irreflexiva, puede parecer que promete conocimiento a nivel reflexivo. De manera característica, espera que las exigencias del conocimiento sean satisfechas solo mediante la reflexión. Sin duda, hay ciertas creencias éticas, sostenidas universalmente, que suelen ser vagas («hay que tener una razón específica para matar a alguien») y que podemos estar seguros

de que sobrevivirán al nivel reflexivo. Pero, a este nivel, están lejos de constituir un cuerpo de conocimiento ético adecuado, y menos aún sistemático, y creo que el resultado de mi exploración anterior de la teoría ética ha mostrado que, tal y como son las cosas, no existe ningún cuerpo de conocimiento de ese tipo. Propondré más adelante, en lo que incumbe al conocimiento proposicional de verdades éticas, que no se trata solo de cómo son las cosas, pues, a un nivel superior de generalidad reflexiva, resulta que no podría haber ningún conocimiento ético de este tipo (o, en el mejor de los casos, podría ser solo parcial).

Si aceptamos que puede haber conocimiento al nivel hipertradicional o irreflexivo; si concedemos la verdad obvia de que es propio de la reflexión perturbar, desestabilizar o reemplazar esos conceptos tradicionales; y si estamos de acuerdo en que, por lo menos en relación con cómo son las cosas, el nivel reflexivo no está en posición de darnos el conocimiento que no poseíamos antes; entonces llegamos a la conclusión, destacadamente antisocrática, de que, en ética, *la reflexión puede destruir el conocimiento.* En el siguiente capítulo, en el que me ocuparé del relativismo, veremos qué significado tiene esta conclusión.

Otra consecuencia, si aceptamos que puede haber conocimiento irreflexivo será la de que no todo conocimiento proposicional es aditivo. No todos los segmentos de conocimiento pueden combinarse en un cuerpo mayor de conocimiento. Puede que tengamos que aceptar esta conclusión después de todo, a causa de otros contextos que implican puntos de vista perspectivistas sobre el mundo. Una parte del mundo físico puede presentar un color a un tipo de observador, y otro a otro tipo, y aún, para otro observador, puede que no tenga color alguno. Llamemos *A, B, C* a esas cualidades que percibe cada tipo de observador. De este modo, un observador avezado de un tipo puede saber que la superficie es *A*, el de otro tipo que es *B*, etc., pero no hay ningún conocimiento de que esta superficie sea *A* y *B* y *C*. Este resultado desaparecerá si lo que significan *A* o *B* o *C* fuera algo relacional, si cuando los observadores dijeran «eso es *A*», quisieran decir «*A* para observadores como nosotros». Es muy dudoso que esta sea la interpretación correcta[13]. Y, si no lo es, la

[13] Véase Wiggins, «Truth, Invention and the Meaning of Life»; y McGinn, *The Subjective View*, págs. 9-10, 119-120.

coherencia de esos segmentos de conocimiento está asegurada a un nivel diferente, cuando las diversas cualidades percibidas guardan relación con la concepción absoluta. Esta relación es también la que explica que las capacidades que producen esos diferentes segmentos de conocimiento sean todas formas de *percepción*. Por supuesto, tenemos buenas razones para creer esto antes de poseer cualquier concepción teórica de ese tipo, y sin duda antes de estar en posesión de sus detalles. Esto se debe a que nuestra experiencia cotidiana, como era de esperar, revela en gran medida lo que somos y cómo nos relacionamos con el mundo, y de esta manera nos encamina hacia la concepción teórica[14].

Hay quien cree que el conocimiento proporcionado por la aplicación de los conceptos éticos es semejante a la percepción, pero ahora podemos ver una asimetría esencial entre el caso de los conceptos éticos y el de la experiencia perspectivista de las cualidades secundarias, tales como los colores. Esta asimetría muestra, además, que la distinción entre lo científico y lo ético tiene implicaciones más amplias. No se reduce a distinguir entre una ciencia idealmente no perspectivista, por un lado, y los conceptos éticos, por el otro. No todos los conceptos perspectivistas son éticos, y existen diferencias significativas entre los conceptos éticos y otros tipos de conceptos perspectivistas, tales como los propios de la percepción sensible.

La diferencia principal es que, en el caso de las cualidades secundarias, lo que explica también justifica; en el caso ético no es así. Las capacidades psicológicas que subyacen a nuestra percepción de las cualidades secundarias del mundo han evolucionado de tal manera que el mundo físico se nos presentará de formas fiables y útiles. Aprender que esas cualidades constituyen nuestro encuentro perceptivo con en el mundo, y que este modo de presentación funciona de cierta manera, no desestabiliza el sistema[15]. La ética presenta una analogía con lo perceptivo en la medida en que hay una convergencia local que subyace a

---

[14] La distinción entre cualidades primarias y secundarias, con independencia de cómo se concrete, es casi tan antigua en la tradición occidental como el uso autoconsciente de un principio de razón suficiente.

[15] He incluido aquí dos frases de mi artículo «Ethics and the Fabric of the World», que aparecerá en Ted Honderich (ed.), *Morality and Objectivity* (Londres, Routledge and Kegan Paul, 1985); se trata de un volumen de seis ensayos en memoria de John Mackie en el que se discuten ideas suyas sobre esas materias; en particular, su idea de que la experiencia perceptiva y la moral son susceptibles de errores comparables. Véase también McGinn, *The Subjective View*, esp. cap. 7.

esos conceptos: los juicios de quienes los usan están guiados por el mundo, como apunté antes. Esto es ciertamente suficiente para refutar las oposiciones más simples entre los hechos y los valores. Ahora bien, si queremos referirnos con esto a algo que tenga una objetividad mayor, todo dependerá de lo que digamos *a continuación*. En el caso de las cualidades secundarias, es la explicación de las percepciones perspectivistas la que nos permite, cuando reflexionamos sobre ellas, ponerlas en relación con las percepciones de otras personas y de otras criaturas, lo cual lo deja todo más o menos como estaba en lo que concierne a nuestros juicios perceptivos. La cuestión es la de si podemos encontrarle una analogía ética. Aquí tenemos que pasar de los juicios locales a su descripción reflexiva o de segundo orden, y entonces la analogía se pierde.

Ante todo, se plantea el problema de qué es una descripción de segundo orden. La *explicación* de esos juicios locales y de las diferencias conceptuales entre sociedades es de suponer que tenga que proceder de las ciencias sociales, dado que se ponen en tela de juicio las diferencias culturales. Tal vez no exista ninguna explicación muy profunda, aunque tampoco tenemos muy claro lo profunda que podría ser una explicación tal. Pero sabemos que no será muy parecida a la explicación de la percepción del color. Las capacidades que invocará son las que nos permiten orientarnos tanto en el mundo físico como en el mundo social, y esto significa, esencialmente, *en un mundo social u otro,* puesto que es cierto, por un lado, que los seres humanos no pueden vivir sin una cultura y, por otro, que hay muchas culturas, que difieren en sus conceptos locales, en las que podemos vivir.

En cualquier caso, una teoría explicativa no basta para hacerse cargo de los problemas relativos a la objetividad planteados por los conceptos éticos locales. En el caso de las cualidades secundarias, la explicación también justifica, porque puede mostrar cómo las percepciones se relacionan con la realidad física y cómo pueden brindar conocimiento de esa realidad, que es lo que pretenden hacer. El problema es si esto nos puede proporcionar un método que nos permita orientarnos en el mundo físico. La explicación teórica muestra cómo lo hace. La reflexión en ética plantea otro tipo de cuestiones. Si nos preguntamos «¿es este un método para orientarnos en el mundo social?», habremos de preguntarnos si se trata de un método para orientarnos en cierto mundo social u otro, y la respuesta tiene que ser, obviamente, afirmativa (a menos que la sociedad fuera extremadamente desordenada, que no es lo que hemos supuesto). La cuestión planteada es, más bien, «¿es este un buen

modo de vivir comparado con otros?», o, por decirlo de otra mane «¿es este el mejor tipo de mundo social?».

Cuando se comprende que estas son las preguntas que se plantea la explicación reflexiva que buscamos resulta que implica consider; ciones *éticas* reflexivas. Estas son las consideraciones que, según cree algunos, habrían de adoptar la forma de una teoría ética. Las consideraciones reflexivas tendrán la tarea de justificar los conceptos locales una vez que estos hayan sido puestos en tela de juicio. Una teoría ética podría incluso, en un sentido débil, proporcionar algunas explicaciones. Podría racionalizar algunas diferencias culturales mostrando por qué un concepto local no es más éticamente apropiado que otros en determinadas circunstancias particulares (podemos recordar aquí las posibilidades y los riesgos del utilitarismo indirecto). Pero, aunque pueda explicar por qué es razonable que las personas tengan esas diversas creencias éticas, una teoría de este tipo no podría explicar por qué las tienen o no. No podría hacer algo que las explicaciones de la percepción sí hacen, a saber: no podría generar una teoría del error adecuada y dar cuenta, en general, de la tendencia de las personas a tener lo que sus propios principios definen como creencias erróneas[16].

Si de todo esto resulta una mayor objetividad, entonces las mismas consideraciones éticas reflexivas tendrían que ser objetivas. Ello nos retorna a la cuestión de si el nivel reflexivo podría generar su propio conocimiento ético. Si esto se entiende como la obtención de conocimiento proposicional de verdades éticas, entonces nos hace falta alguna explicación de lo que significa «rastrear la verdad». La idea de que nuestras creencias pueden rastrear la verdad a este nivel tiene que implicar, por lo menos, que un grupo de investigadores podrían llegar a converger de manera racional, razonable y sin condicionamientos sobre un conjunto determinado de conclusiones éticas. ¿Qué esperanza hay de que pueda darse este proceso? No quiero decir que acaezca efectivamente, sino que nos formemos una imagen coherente de cómo podría acaecer. Si se entiende que es un proceso de convergencia relativa a un

---

[16] Esta dificultad de hallar una teoría del error adecuada acecha a cualquier teoría de la ética que se centre en la noción de verdad ética. Cuando lo ético adopta la forma específica de *la moral,* se vincula con una deformación particular, el *moralismo.* La insistencia en que cierta persona está equivocada, sin la comprensión de cómo ha llegado al error, tiene como consecuencia que el comentarista actúe de una forma completamente ajena a esa persona y termine sermoneándola.

cuerpo de verdades éticas cuya causa y explicación reside en el hecho de que son verdades —esta sería la analogía estricta con la objetividad científica—, entonces no tengo ninguna esperanza depositada en ello. En concreto, no cabe esperar que, a este nivel, el mundo pueda guiarnos de la manera en que hemos evocado en el caso de los conceptos éticos densos. Los debates a nivel reflexivo, si son lo bastante ambiciosos como para considerar toda la experiencia ética y llegar a la verdad sobre lo ético, usarán necesariamente los conceptos éticos más generales y abstractos, tales como «correcto», con respecto a los cuales el mundo no ejerce ningún tipo de guía (que es la razón por la que los eligió el prescriptivismo en su intento de hallar un elemento evaluativo puro del que poder *cercenar* la guía del mundo).

No alcanzo a encontrar una teoría del conocimiento convincente de la convergencia del pensamiento ético reflexivo sobre la realidad ética que tenga visos de permitir ni siquiera una analogía remota con la ciencia. Tampoco existe ninguna analogía convincente con las matemáticas, un caso en el que la noción de una realidad independiente es, cuanto menos, problemática. Más allá de las razones mencionadas en el capítulo 6, lo importante es que toda sección no contradictoria de las matemáticas es parte de las matemáticas, aunque esto cabe dejarlo de lado por trivial, poco esclarecedor o inútil. Pero no toda estructura no contradictoria éticamente reflexiva puede formar parte de la ética, puesto que los cuerpos de pensamiento ético pueden entrar en conflicto entre sí de modo que no solo carezcan de los tipos de explicación susceptibles de constituir una teoría del error que sea creíble, sino que tampoco dispongan de muchas explicaciones creíbles de otros tipos.

No creo, pues, que podamos concebir la reflexión como un proceso que venga a sustituir el conocimiento por creencias adquiridas mediante una práctica irreflexiva. Tenemos que rechazar el punto de vista objetivista de la vida ética, que la concibe como una búsqueda, en ese sentido, de la verdad ética. Pero esto no descarta todas las formas de objetivismo. Nos queda aún el proyecto de tratar de dar una base o un fundamento objetivos a la vida ética. Para ello, hemos de mirar en la dirección de las ideas sobre la naturaleza humana que examinamos en el capítulo 3. Esas ideas se liberarían ahora de la exigencia socrática de tener que proporcionar a *cada* persona una razón para preferir llevar una vida ética. Para nuestros propósitos actuales, sería bastante signifi-

cativo que tales consideraciones pudieran darnos un esquema de vida ética que fuera la mejor vida ética, la más satisfactoria para los seres humanos en general. La pregunta que hemos de responder es: admitiendo que los seres humanos necesitamos compartir un mundo social, ¿hay algo que habríamos de conocer con respecto a nuestras necesidades y motivaciones fundamentales que nos revelara que ese mundo sería el mejor?

Dudo de que podamos obtener una respuesta muy satisfactoria. Es probable que consideraciones semejantes basten para determinar las opciones éticas incluso en una situación social dada (tenemos que recordar que tomamos la situación en sí misma, en parte, en función de las opciones éticas que podemos detectar). Cualquier vida ética impondrá limitaciones a cosas tales como matar, dañar y mentir, pero estas limitaciones pueden adoptar formas muy diferentes. De nuevo, con respecto a las virtudes, que constituyen el campo más natural y prometedor para este tipo de investigación, nos basta con comparar el catálogo de virtudes de Aristóteles con cualquier otro que podamos confeccionar hoy en día para apreciar cómo las imágenes de una vida humana apropiada pueden diferir en su espíritu y en las acciones y las instituciones que recomiendan. También albergamos la idea de que existen muchas y muy diversas formas de excelencia humana que no encajarán, todas, en un conjunto armonioso, de modo que cualquier perspectiva ética determinada representará una especialización de las posibilidades humanas. Esta idea está profundamente enraizada en cualquier concepción naturalista o histórica de la naturaleza humana —esto es, sensata—, y me parece difícil creer que pueda ser superada por una investigación objetiva o que pueda resultar que los seres humanos tenemos una naturaleza mucho más determinada de lo que sugiere lo que ya sabemos; es decir, una naturaleza que demande siempre una vida de un tipo particular.

A mi entender, es muy poco probable que el proyecto de dar a la vida ética una base objetiva y determinada mediante consideraciones sobre la naturaleza humana tenga éxito. Pero, por lo menos, es un proyecto comprehensivo que creo que representa la única forma inteligible de objetividad ética en el nivel reflexivo. Por ello, merece la pena que nos preguntemos en qué consistiría su éxito. Cabe señalar, en primer lugar, cómo habrían de ser los seres humanos que fueran los sujetos primarios de nuestra ética, dado que esta extraería sus conclusiones a partir de su naturaleza. En este punto, el proyecto va de la mano con el

contractualismo, al entender que los demás animales están fuera de la circunscripción primaria de la ética y son solo, a lo sumo, sus beneficiarios, mientras que tiene menos expectativas que el contractualismo con respecto a nuestras relaciones con extraterrestres, que se relacionarían con el proyecto solo a través de reglas de limitación recíproca parecidas a las de un tratado de no agresión.

Que el proyecto pueda llegar a buen puerto no será solo cuestión de concordar respecto a una teoría de la naturaleza humana. La convergencia misma se daría en parte en la ciencia social y en la psicología, pero lo importante sería una convergencia a la que las conclusiones científicas contribuyeran solo en parte. Por otro lado, no habría convergencia directa con respecto a las verdades éticas, como en el modelo objetivista. Se podría decir que una creencia ética sería un objeto de conocimiento de pleno derecho a nivel reflexivo, con el resultado de que un determinado tipo de vida sería el mejor para los seres humanos. Pero esto no conduciría directamente a otras verdades éticas. La razón de ello, dicho sucintamente, es que la excelencia o el carácter satisfactorio de una vida no guarda la misma relación con las creencias implicadas en esa vida que la relación entre premisas y conclusión. Por el contrario, la vida excelente de un agente está caracterizada por el hecho de que él *tiene* esas creencias, y la mayoría de ellas no se referirán ni a las disposiciones ni a la vida de ese agente ni tampoco a las disposiciones de otras personas, sino al mundo social. Esa vida entrañará, por ejemplo, que el agente utilice ciertos conceptos densos y no otros. La reflexión sobre la excelencia de una vida no basta para establecer la verdad de los juicios que usan esos conceptos, ni de los demás juicios éticos del agente. Lo que muestra es que hay buenas razones (asumiendo el compromiso con una vida ética) para vivir una vida que implique esos conceptos y esas creencias[17].

La convergencia que señalaría el éxito de este proyecto sería una convergencia relativa a la razón práctica, gracias a la cual las personas lograrían llevar el mejor género de vida y tendrían los deseos que se corresponden con esa vida; la convergencia en la creencia ética sería, en gran medida, parte y consecuencia de ese proceso. Una creencia ética muy general constituiría efectivamente un objeto de conocimiento a

[17] Esta conclusión conecta con la conclusión del capítulo 3, según la cual todos los valores descansan en disposiciones de carácter. Véase también el Epílogo a este libro.

ese nivel. Muchos juicios éticos particulares, incluyendo los conceptos densos que uno favorece, podría saberse que son verdaderos, pero —como he defendido— a menudo se sabe de todos modos que los juicios de este tipo son verdaderos, incluso cuando acaecen, como siempre ha sucedido, en una vida sin fundamento objetivo. El fundamento objetivo no supondría que los juicios que usan esos conceptos fueran verdaderos o pudiesen conocerse; esto ya era así. El fundamento objetivo nos permitiría reconocer que algunos de estos conceptos son los mejores o más apropiados conceptos densos que podemos usar. Entre los dos extremos que conforman una proposición muy general y las múltiples proposiciones concretas, otras creencias éticas serían verdaderas solo en el sentido oblicuo de que serían las creencias que nos ayudarían a orientarnos en el mundo social que —según este programa optimista— se ha mostrado como el mejor mundo social para los seres humanos.

Esta sería una estructura muy diferente a la de la objetividad de la ciencia. Habría una diferencia radical entre la ética y la ciencia, incluso si la ética fuera objetiva en el único sentido en que puede serlo de manera inteligible. Sin embargo, esto no significa que exista una distinción clara entre (cualquier) hecho y (cualquier) valor; no significa que no haya conocimiento ético. Lo hay, y en un pasado menos reflexivo lo hubo mucho más.

Los problemas que he analizado aquí no son simples cuestiones hipotéticas relativas a si la ética puede ser objetiva y, en su caso, a cómo lo sería. Son problemas referidos a la naturaleza del pensamiento ético, al modo en que este puede comprender su propia naturaleza y a la medida en que puede parecer ser, de forma coherente, lo que realmente es. Esos son problemas serios para cualquier planteamiento, y lo serían incluso si el pensamiento ético resultara ser objetivo en el único sentido que es inteligible. Nos ocuparemos de ellos más detenidamente cuando los examinemos desde un ángulo diferente, el del relativismo.

Capítulo 9

Relativismo y reflexión

Si, al reflexionar sobre cierto tipo de desacuerdos, llegamos a la conclusión de que no es posible resolverlos objetivamente, podemos reaccionar adoptando alguna forma de relativismo. El relativismo no es algo peculiar a la ética; lo encontramos en muchos ámbitos, incluso en filosofía de la ciencia. Su meta es afrontar puntos de vista, perspectivas o creencias que están aparentemente en conflicto y tratarlos de tal manera que dejen de estarlo: cada una resulta ser aceptable a su manera. El problema es encontrar un procedimiento que lo permita; en concreto, encontrar el momento y el lugar que le corresponde a cada creencia o perspectiva.

El método más simple, y el que es relativista en el sentido más preciso del término, es interpretar las tesis iniciales como si cada una introdujese una relación con un elemento diferente. El pensador griego Protágoras, que suele ser reconocido como el primer relativista, partió del conflicto entre las apariencias sensoriales —como cuando a mí me parece que el viento es frío y otra persona lo encuentra cálido— y afirmó que no había una respuesta a la pregunta de si realmente el viento era «en sí mismo» frío o cálido. El hecho era que es frío para mí y cálido para la otra persona. Este es el planteamiento relacional del tema que mencioné en el capítulo anterior como una manera, aunque no la más convincente, de tratar las variaciones en la percepción de cualidades secundarias, pero no entre individuos (como en el caso original de Protágoras) sino entre diferentes tipos de perceptores.

El objetivo del relativismo es *disipar un conflicto explicándolo,* y esto conlleva dos tareas: tiene que decirnos por qué, en realidad, no hay ningún conflicto, y también por qué parecía que lo había. El relativismo estrictamente relacional cumple la primera tarea resueltamente, al hallar en los dos enunciados una forma lógica que los hace directamente compatibles, de modo que aceptarlos ambos no supone ningún problema. Tiene menos éxito en la segunda tarea, porque, cuanto más convincente resulta afirmar que los enunciados son realmente relacionales, más desconcertante es que las personas hayan creído que había un conflicto entre ellos. El relativismo relacional introduce una estructura que los hace claramente compatibles y, por ello, ha de explicar qué la enmascaraba. Puede ser útil aproximarnos al relativismo desde la otra dirección y preguntarnos qué sucede si empezamos concediendo que dos creencias o dos perspectivas entran realmente en conflicto y se excluyen mutuamente de manera genuina. El problema será, pues, hallar un sentido en el que cada uno pueda ser aún aceptable en su lugar.

Una idea que nos hace plantearnos la cuestión del relativismo de una manera más general es la de la *inconmensurabilidad.* Algunos filósofos de la ciencia sostienen que las teorías científicas pueden ser inconmensurables entre sí porque difieren en los conceptos que usan, en las referencias que dan a diversos términos y en lo que consideran evidencia. No es que esas teorías se contradigan directamente, pero sí se excluyen entre sí. Si no lo hicieran, no habría ninguna dificultad en combinarlas, al igual que podemos combinar la topografía de lugares distintos. Pero no pueden combinarse, y fue esto lo que inició el debate. Quienes defienden una de las teorías tratan de encontrar razones para rechazar la otra; una de ellas puede expulsar a la otra en el curso de la historia de la ciencia. ¿Cómo es esto posible? Algunos filósofos de la ciencia radicales dirán que no podemos combinar las dos teorías simplemente porque no podemos combinar la aceptación de ambas teorías: las actividades investigadoras características de cada teoría, la orientación de la atención propia de cada una, etc., no pueden combinarse. No podemos funcionar con ambas teorías conjuntamente.

Esta caracterización de las teorías científicas rivales hace que parezcan dos culturas o formas de vida distintas. Como concepción de la ciencia, me parece de una exageración desmesurada, pero un relato de este tipo puede resultar apropiado para las que son realmente culturas o formas de vida diferentes, como la de la sociedad hipertradicional del capítulo precedente. La perspectiva de una de esas sociedades podría ser

inconmensurable, en gran medida, con la de otra, pero aún se excluirían mutuamente. El conflicto residiría en la forma de vida que cada una de ellas implica.

Si dos culturas o perspectivas o formas de vida se excluyen mutuamente, ¿hay margen para el relativismo? No inmediatamente. Confrontado con un modo de vida alternativo, alguien que tiene determinadas disposiciones y expectativas como miembro de una cultura normalmente no querrá hacer lo que se hace en la otra cultura. Además, lo que confiere un carácter ético a sus reacciones es, en parte, que las tiene tan interiorizadas que, en ciertos casos, no son solo expresión de mera reticencia sino de rechazo. Para que el rechazo sea apropiado no es necesario que las partes conceptualicen del mismo modo las acciones en cuestión y, dada la situación que estamos planteando, no lo harán. Así pues, cuando los miembros de una cultura que no admite el sacrificio humano se encuentran con miembros de una que sí lo hace, conceptualizan la matanza ritual de forma diferente, pero esto no significa que el primer grupo, si se horroriza, esté cayendo en un malentendido antropológico. Como dirían sus miembros, se trata de matar deliberadamente a un cautivo, lo cual basta para que viertan sobre este acto sus sentimientos éticos de hostilidad. (Pero no se sigue que tengan que censurarse mutuamente; ese es otro tema).

El hecho de que los miembros de una cultura estén expuestos a las conductas de los miembros de otra cultura no hace que sus disposiciones y reacciones cambien o dejen de ser apropiadas. En todo caso, resulta artificial tratar estos asuntos como si en ellos siempre estuvieran implicadas dos culturas claramente delimitadas. Una cultura completamente individualizada es algo raro, en el mejor de los casos. Culturas, subculturas, fragmentos de culturas, tienen encuentros constantes, así como intercambian y modifican sus prácticas y actitudes. Las prácticas sociales nunca podrían presentarse con un certificado que dijera que pertenecen a una cultura genuinamente diferente que garantizase su inmunidad ante los juicios y las reacciones ajenos.

Queda, así, descartado el relativismo instantáneo. Por razones semejantes, el relativismo relacional estricto aplicado a la ética también queda descartado. Ha tenido defensores hábiles[1], pero es implausible suponer

---

[1] Véase, en concreto, Gilbert Harman, «Moral Relativism Defended», *Philosophical Review*, 84 (1975); reimpreso en Michael Krausz y Jack W. Meiland (eds.), *Relativism, Cognitive and Moral* (Notre Dame, University of Notre Dame Press, 1982), que constituye una compilación útil sobre este tema.

que las concepciones éticas de lo que está bien y lo que está mal tienen una relatividad lógicamente inherente a una sociedad dada. Tomemos de nuevo el caso de la sociedad hipertradicional y supongamos que tiene ciertas reglas que se expresan en términos de «bueno» y «malo». Cuando entra en contacto con otra cultura y se la invita a reflexionar, no puede descubrir repentinamente que hay una relativización implícita que se oculta en su lenguaje. Siempre será, puede decirse, demasiado pronto o demasiado tarde para eso. Es demasiado pronto si no han reflexionado o pensado nunca sobre una alternativa al «nosotros». (Una de las preguntas planteadas en el capítulo 7 se aplica a esto: ¿Cómo esta reflexión puede formar parte de su lenguaje?) Es demasiado tarde si confrontan la situación nueva; eso les exige ver más allá de sus reglas y de sus prácticas.

Parece ahora como si pudiéramos descartar el relativismo por completo. El hecho de que las personas puedan y tengan que reaccionar cuando se ven confrontadas con otra cultura, y que lo hagan aplicando sus nociones —también reflexionando sobre ellas— parece mostrar que el pensamiento ético de una cultura dada puede siempre sobrepasar sus límites. Es importante advertir que esto se refiere al contenido o a las aspiraciones del pensamiento ético, no a su objetividad. Incluso si no hay modo alguno en el que las creencias éticas divergentes puedan converger por medio de una indagación independiente o de la argumentación racional, esto no implica el relativismo. Cada perspectiva puede aún enunciar tesis que pretendan aplicarse al mundo en su conjunto, y no solo a la parte de él que constituye su «propio» mundo.

Sin embargo, aunque es cierto que la falta de objetividad no implica ninguna actitud relativista, parece haber algo vacuo e insensible en el hecho de conformarse con esta verdad. Si somos *conscientes* de la falta de objetividad, ¿no habría de afectar esto al modo en que concebimos la aplicación o la extensión de nuestra perspectiva ética? Y si es así, ¿cómo? Esta consciencia no puede simplemente apagar nuestras reacciones éticas cuando nos vemos confrontados con otro grupo, y no hay ninguna razón por la que haya de ser así. Algunas personas han pensado que sí habría de ser así, pues creen que una concepción propiamente relativista exige que tengamos la misma buena disposición hacia las creencias éticas de todo el mundo. Este punto de vista es gravemente erróneo, pues concibe el relativismo como una moral no relativista de la tolerancia universal[2]. Pero,

---

[2] Examino el relativismo vulgar, como lo he llamado, en mi *Introducción a la ética*.

aunque suponga incurrir en una confusión, no cabe duda de que es una reacción a algo. Si llegamos a percatarnos de la variabilidad ética y de las diversas explicaciones que esta puede recibir, es inconcebible que esta consciencia lo deje todo como estaba y no afecte a nuestro mismo pensamiento ético. No cabe duda de que podemos continuar diciendo que nosotros estamos en lo cierto y que son los demás quienes se equivocan (es decir, según el punto de vista no objetivista, afirmando nuestros valores y rechazando los suyos), pero parece una respuesta claramente inadecuada si hemos llegado a este estadio reflexivo. ¿Qué otras posibilidades hay? Para responder a esto, volvemos una vez más a la cuestión del relativismo. Tradicionalmente, la cuestión ha sido la de si *tenemos que* pensar de un modo relativista por razones conceptuales o lógicas o si esto es imposible. Pero lo que verdaderamente habríamos de preguntarnos es en qué medida podemos pensar coherentemente de este modo y hasta qué punto esto da una respuesta más adecuada a la reflexión.

Todas las ideas que hemos considerado hasta el momento presuponen la existencia de una distinción fundamental entre la perspectiva de un grupo y la de todos los demás. El relativista piensa que los juicios de un grupo se aplican solo a ese grupo, y la parte contraria piensa que los juicios de cualquier grupo tienen que aplicarse a todo el mundo. Ambos están equivocados. Para hacerse cargo de las preocupaciones del relativista no basta con trazar una línea divisoria entre nosotros y los demás. No es cuestión de trazar ninguna línea, sino de reconocer que los demás se encuentran a diferentes grados de distancia de nosotros. Tenemos que ver también que nuestras reacciones ante los otros grupos y nuestras relaciones con ellos forman parte de nuestra vida ética, y hemos de concebir esas reacciones de un modo más realista, en términos de las prácticas y los sentimientos que ayudan a estructurar nuestra vida. Algunos desacuerdos y divergencias tienen más importancia que otros. Pero, sobre todo, hay que ver si el contraste entre nuestra perspectiva y cualquier otra es tal que marca la diferencia con respecto a la cuestión del tipo de vida que vivirá un grupo u otro.

Hay que distinguir entre confrontaciones *reales* y *nocionales*[3]. Una confrontación real entre dos perspectivas divergentes se da cuando, en

---

[3] He presentado esta idea, más detalladamente, en «The Truth in Relativism», *Proceedings of the Aristotelian Society*, 75 (1974-1975), reimpreso en mi libro *La fortuna*

un momento dado, hay un grupo de personas para quienes cada una de las perspectivas es una opción real. Por el contrario, una confrontación nocional se da cuando ciertas personas saben de dos perspectivas divergentes, pero al menos una de ellas no representa una opción real. La idea de «opción real» es, en gran parte, pero no del todo, una noción social. Una perspectiva es una opción real para un grupo si, de hecho, es ya su perspectiva o si podrían adoptarla; y podrían adoptarla si pudieran vivir en su seno, según sus circunstancias históricas presentes, manteniendo el contacto con la realidad, esto es, sin caer en el autoengaño, etc. La medida en la que pueden hacer esto depende de qué características de su situación social presente se asume que permanecerán constantes si cambian a la otra perspectiva. Una cosa podría volverse posible para ciertas personas si su situación cambiara, y la cuestión de si es una opción real para ellas implica saber si se podría cambiar su situación. Las personas pueden estar equivocadas sobre estas cuestiones. Pueden pensar que una perspectiva es una opción real cuando no lo es, porque estén mal informadas o sean muy optimistas o caigan prisioneras de sus fantasías, y este puede ser un error no solo personal, sino también social o político. Por el otro lado, puede que no se hayan percatado de lo que el tránsito a la nueva perspectiva podría ofrecerles.

Muchas perspectivas que han tenido los seres humanos no son opciones reales para nosotros ahora. La vida de un cacique de la Edad del Bronce o de un samurái medieval no son opciones reales para nosotros: no hay manera de que podamos vivirlas. Con esto no niego que la reflexión sobre esos sistemas de valores pueda inspirar ciertos pensamientos relevantes para la vida moderna, sino que no es posible adoptar esas perspectivas. Ni siquiera proyectos utópicos ideados por un pequeño grupo de entusiastas podrían reproducir *esa* vida. Más aún, el proyecto de resucitarlos a escala de una sociedad en el contexto de la vida industrial moderna conllevaría una vasta ilusión social. Las esperanzas de eliminar, en conjunto, las condiciones de la vida industrial moderna suponen algo más: otra imposibilidad, aunque distinta.

Es importante caer en la cuenta de que esas opciones pueden estar relacionadas entre sí de forma asimétrica. Algunas versiones de la vida

---

moral («La verdad en el relativismo»), de donde he adaptado unas pocas frases en lo que sigue. Como veremos, no quiero seguir defendiendo sin matices lo que dije allí: que un punto de vista relativista, definido en esos términos, es correcto para las perspectivas éticas *(Moral Luck,* pág. 142).

tecnológica moderna se han convertido en una opción real para miembros de sociedades tradicionales que sobreviven actualmente, pero su vida no es una opción real para nosotros, a pesar del entusiasmo nostálgico de algunos. Las teorías de las que disponemos sobre la naturaleza y alcance de tales asimetrías afectan a nuestra visión de las posibilidades de la acción social y política radical.

Puede entenderse que la concepción relativista de un cierto tipo de perspectiva afirma que, a este tipo de perspectiva, solo puede aplicarse el lenguaje de la evaluación (bueno, malo, correcto, incorrecto, etc.) en las confrontaciones reales. En las nocionales, este tipo de evaluación se considera inapropiado y no constituye ninguna clase de juicio. Cuando se rechaza el relativismo en un área dada, esto no significa que no haya ninguna confrontación nocional. La confrontación entre la teoría del flogisto y cualquier teoría contemporánea de la combustión es, sin duda, nocional, y la teoría del flogisto no es hoy una opción real; pero la concepción no relativista de tales teorías permite decir algo como evaluación de la teoría del flogisto: que es falsa. No es solo que tratar de vivir la vida de un teórico del flogisto convencido en el mundo académico contemporáneo es una empresa tan incoherente como la de querer vivir la vida de un caballero teutón en el Núremberg de los años treinta del siglo xx; la teoría del flogisto no es una opción real porque no cuadra con un montón de cosas que sabemos que son verdaderas.

Al relativismo entendido de esta manera lo denominaré *relativismo de la distancia*. Hay margen para él en una perspectiva ética reflexiva. La distancia que convierte en nocional la confrontación, y que hace posible este tipo de relativismo, puede adoptar diversas direcciones. A veces, es cuestión de lo que está en otro lugar, que es cuando el relativismo se aplica a lo exótico. Es natural aplicarlo al pasado más lejano. Y puede aplicarse también al futuro; me ocuparé de ello en la parte final de este capítulo.

Al presentar este tipo de relativismo he mencionado perspectivas éticas en vez de prácticas particulares, y el relativismo tiene que aplicarse a sistemas o cuerpos de creencias y actitudes a gran escala. Si hemos de tomarnos en serio la suspensión relativista del juicio ético, tenemos que concebir la sociedad en cuestión como un todo. Hay algunos conceptos éticos que podemos aplicar a las personas y a sus acciones —los conceptos de virtud y vicio, por ejemplo—, incluso si no hay una confrontación real entre la perspectiva de la sociedad en la que vivían y la nuestra. Esto implica considerar a las personas haciendo abstracción

de las prácticas sociales en las que vivían, y por ello, a menudo, no los consideraremos de forma realista. Un caso especial es la figura histórica de quien ha sido un criminal o un disidente, que ni nosotros ni sus contemporáneos consideramos que viviera completamente con arreglo a los valores locales. En este caso, en principio cabe una comprensión concreta del disidente y de la sociedad, aunque raramente se da[4].

Si nos tomamos en serio la suspensión relativista de la evaluación, tenemos que pensar en la sociedad misma de modo realista y concreto. Muchas teorías éticas que nos hablan del pasado o de mundos exóticos tienen poco que ver con la realidad de esas épocas y esos lugares. Son fantasías, que sirven a algunos de los mismos fines éticos que los cuentos de hadas y, si ofenden a alguien, es a la concepción realista de la vida y de las posibilidades humanas, y no a un relativismo seriamente planteado. No se interesan realmente por otras sociedades, sino que solo las usan como fuente de símbolos y de aspiraciones[5].

El relativismo referido solo a la distancia espacial no es de ningún interés o aplicación en el mundo moderno. Hoy en día todas las confrontaciones entre culturas tienen que ser confrontaciones reales, y la existencia de sociedades tradicionales exóticas plantea cuestiones bastante distintas y difíciles de resolver, tal como la de si el resto del mundo puede o habría de usar el poder para preservarlas, como si fuesen especies en peligro de extinción; los antropólogos y otros investigadores de campo se encuentran a sí mismos ejerciendo el papel de guardianes del coto[6]. Los pensamientos sobre el pasado y el futuro plantean diferentes problemas porque somos el efecto de uno y la causa del otro. Además, el pasado y la comprensión que tenemos de él guardan una relación especial con la reflexividad que originó estos problemas. Considero que el mundo moderno está marcado por un nivel particular de reflexividad, y —aunque esto ya lo expresó Hegel y esos debates en buena medida empezaron con él— el conjunto de marcos explicativos en los que podemos situar nuestra cultura y la de los demás es mucho mayor hoy en día.

---

[4] Para un ejemplo del surgimiento de una leyenda —o de diferentes leyendas—, véase J. C. Holt, *Robin Hood* (Londres, Thames and Hudson, 1982).

[5] Una contribución muy interesante a esta temática es Bernard Smith, *European Vision and the South Pacific* (Nueva York, Oxford University Press, 1969).

[6] Por la misma razón, cuando imaginamos el futuro ya no pensamos en pueblos exóticos sino en extraterrestres. Dado que no pueden ofrecer ninguna resistencia concreta a todas las fantasías más primarias, los resultados son pobres, patéticos y repulsivos.

El crecimiento de la consciencia reflexiva no ha sido uniforme ni siempre positivo. Aún menos hemos de creer que hasta cierto punto hubo en el mundo occidental una vida comunitaria integrada, concreta y familiar, que saltó por los aires a causa de algo que, según gustos, cabe identificar con la guerra de 1914, la Revolución Industrial, Galileo, la Reforma u otros acontecimientos anteriores. Esas diversas versiones de la Caída son todas igualmente míticas y expresan la nostalgia de un estado de absoluta identificación con el ambiente, la nostalgia de un recuerdo brumoso. No tenemos por qué aceptar el mito para dar por establecidas dos cosas: que la necesidad de una comprensión reflexiva de nuestra sociedad y de nuestras actividades es más profunda y más extensa de lo que lo ha sido nunca antes; y que los conceptos éticos más densos son menos corrientes en la sociedad moderna que en las sociedades tradicionales, incluso aunque su uso en esas sociedades no garantizara, como pretende el mito, la identidad comunitaria, la ausencia de conflicto y un sentimiento de plenitud[7].

No hay una ruta para volver atrás y renunciar a la reflexividad. No quiero decir que nada pueda llevarnos en esa dirección; tanto personal como socialmente, muchas cosas pueden hacerlo. Pero ya no hay un *camino* de vuelta que nos permita escapar conscientemente de ella. Incluso en lo individual, aunque podamos embarcarnos conscientemente en la tarea de reducir la reflexión, no podemos tratar de realizar conscientemente esa tarea; y, si conseguimos llevarla a cabo, tendremos que olvidarlo. Pero a nivel social habrá personas que no quieran tratar de realizar esta tarea, que intentarán hacer que los demás no lo olviden. Este fenómeno de autoconsciencia, junto con las instituciones y los procesos que la apoyan, constituyen una de las razones por las que las formas de vida del pasado no son una opción real para el presente, y por la que los intentos de volver al pasado suelen producir resultados grotescos a pequeña escala y espantosos a gran escala. Esto puede verse, sobre todo, en los proyectos reaccionarios que pretenden recrear sociedades jerárquicas del pasado supuestamente dichosas. Estos proyectos se enfrentan, en cualquier caso, a la crítica de que sus concepciones del pasado son fantasiosas; pero, incluso si pueden existir jerarquías dichosas, no tendrán más encanto para nosotros que el de su inocencia y el

---

[7] Alasdair MacIntyre, en *Tras la virtud*, está interesado en este tema, aunque muestra cierta debilidad por el mito.

desconocimiento de su propia naturaleza. Esto no puede recrearse, pues habría que tomar medidas para impedir que las personas plantearan cuestiones que, a estas alturas, ya están aquí para ser planteadas[8].

Pero, una vez las preguntas están ahí para ser planteadas, ¿no hemos de plantearlas con respecto a esas sociedades tal y como existieron en el pasado?, ¿acaso podríamos no hacerlo? En concreto, ¿no podemos preguntarnos si esas sociedades, aunque no se percataran, fueron injustas? ¿Puede el relativismo de la distancia protegerlas de esta pregunta? Hay una cuestión que resultará útil que nos formulemos desde el principio. Al ser menos reflexivas y autoconscientes que la sociedad moderna, ¿qué es lo que esas sociedades ignoraban?

Es tentador afirmar que no conocían alternativas a su organización social y pensaron que su orden social era necesario. Esto puede ser cierto con respecto a ciertas sociedades tradicionales, aisladas y analfabetas, pero muchas sociedades jerárquicas sofisticadas, las de la Europa medieval, por ejemplo, sin duda supieron que había alternativas al orden establecido, en tanto que sabían que, en la misma época y en otros lugares, los seres humanos se habían organizado en otros tipos de sociedades. Lo que no sabían, tendremos que decir, es que había una alternativa *para ellos*. Pero, en tal caso, no es evidente que hubiera una alternativa para ellos. Nos hace falta concebir más claramente *qué hubiera podido significar* decir (en un sentido, al menos, que sea interesante) que podrían haber tenido una organización social diferente[9], e ideas aún más robustas de la libertad, para poder decir que eran capaces de conseguirla. Quizá no se equivocaran al pensar que su orden social era necesario para ellos. Lo que no podemos aceptar es el modo en el que concibieron que era (religiosa o metafísicamente) necesario. En nuestra opinión, su error reside en los mitos que legitimaban sus jerarquías. Consideramos que la concepción que nosotros tenemos de nuestra sociedad y de nosotros mismos es más naturalista que la concepción que esas sociedades tenían de sí mismas. Esta concepción naturalista de la

---

[8] «... cuando el velo empieza a rasgarse, no admite reparación. La ignorancia tiene un carácter peculiar: una vez que se disipa es imposible restablecerla. No es inicialmente una cosa en sí, sino únicamente la ausencia de conocimiento, y aunque puede *mantenerse* al hombre en la ignorancia, no se lo puede *hacer* ignorante». Thomas Paine, *Los derechos del hombre*, parte primera [trad. de F. Santos Fontela (Madrid, Alianza, 1984)].

[9] Esto lo analiza Geoffrey Hawthorn en *Plausible Worlds* (Cambridge University Press, 1991).

sociedad, articulada por Hobbes y Spinoza en los albores del mundo moderno, representa una de las maneras en que, según la célebre expresión de Max Webber, el mundo ha sido *entzaubert,* ha sido desencantado. (En particular, los intentos actuales de las fuerzas islámicas por revertir ese proceso —si sus intentos son realmente eso— no muestran que el proceso sea local o reversible, sino solo que puede generar desesperación).

Las legitimaciones de la jerarquía ofrecidas por las sociedades del pasado, y cómo las entendemos hoy en día, son relevantes para lo que digamos sobre la justicia o injusticia de esas sociedades. Esto es importante para el relativismo de la distancia. «Justo» e «injusto» son términos fundamentales que pueden aplicarse a las sociedades como un todo, así como a sociedades concebidas de forma concreta y realista, siquiera en principio. Además, una evaluación en términos de justicia puede realizarse, más obviamente que otras, sin entrar en la inútil cuestión de a quién hay que censurar. La combinación de esas características hace de la justicia social un caso especial en relación con el relativismo. La justicia y la injusticia son, evidentemente, nociones éticas que resultan aplicables a las sociedades del pasado como un todo, incluso cuando las comprendemos bastante bien.

Es defendible una concepción relativista de la justicia. Cuando pensamos históricamente, tendemos a creer que las concepciones modernas de la justicia social, la igualdad de derechos, por ejemplo, no son aplicables a las sociedades jerárquicas del pasado. El hecho evidente de que esas sociedades no satisfarían las condiciones de Rawls, citadas en el capítulo 5, no parece relevante ni para esas sociedades ni para los méritos de los criterios que Rawls ha propuesto específicamente para las sociedades modernas[10]. Aun así, hay fuertes presiones para que las justicias o injusticias de las sociedades del pasado no se desvanezcan sin más en el relativismo de la distancia. Aunque nos opongamos a aplicarles determinadas ideas modernas, no hay duda de que en esas sociedades se usaron ciertas concepciones de la justicia, y no es un juego de palabras ni un error lingüístico llamarlas así. Podemos considerar que ciertas concepciones modernas de la justicia social son aplicaciones más radicales —erróneas, dirían los conservadores— de ideas que han exis-

[10] En su *Teoría de la justicia,* Rawls parece no haber considerado la cuestión en términos específicamente históricos. En trabajos más recientes, ha subrayado la idea de que esta concepción de la justicia está especialmente adaptada a las sociedades modernas.

tido en otras partes y dieron forma a otras sociedades; igualmente, podemos dar un uso ético a las continuidades históricas en sentido inverso. Las concepciones precedentes, en cierta forma, siguen entre nosotros. Sabemos, o habríamos de saber, que no hay manera de volver atrás y que las legitimaciones de la estructura jerárquica de las sociedades del pasado ya no son una opción. Pero, si los radicales pueden identificar determinadas concepciones modernas más igualitarias como un legado de concepciones de la justicia del pasado, los conservadores también pueden tratar de hallar algún análogo menos igualitarista de las concepciones antiguas para hoy en día, liberado de las legitimaciones del pasado, que (a menos que sean reaccionarios oscurantistas) pueden ver, como vemos todos, que ya no funcionan.

Es muy importante para el pensamiento ético concretar en qué sentido las legitimaciones del pasado se consideran desacreditadas. Tanto el aumento de la reflexión como la concepción naturalista de la sociedad conducen a su descrédito intelectual; son así explicadas o comprendidas, pero no de la manera en que ellas hubieran deseado. No obstante, algunas explicaciones pueden hacer que queden desacreditadas éticamente. La teoría crítica[11] insta acertadamente a plantear cuestiones concretas sobre su explicación, preguntando si la aceptación de la legitimación no puede haber sido solo un efecto del poder al que se suponía que legitimaba. Potencialmente, este es un argumento ético, no solo explicativo. Tales argumentos contribuyen a mantener vivas las preguntas sobre la justicia de sociedades del pasado dentro de los límites del pensamiento ético moderno, y a generar dudas sobre el relativismo de la distancia.

Hay aún mucho más que decir sobre estos temas[12]. Puede ser que las consideraciones relativas a la justicia sean un elemento central del

---

[11] Son obras secundarias útiles: Raymond Geuss, *The Idea of a Critical Theory* (Nueva York, Cambridge University Press, 1981); Martin Jay, *The Dialectical Imagination* (Boston, Little, Brown, 1973) [*La imaginación dialéctica,* trad. de J. C. Curutchet (Madrid, Taurus, 1989)], así como *Adorno* (Cambridge, Harvard University Pres, 1984) [*Adorno,* trad. M. Pascual Morales (Madrid, Siglo XXI, 1998)]. La teoría crítica ha pagado, particularmente en la pasada década, una pena justa por su estilo de pensamiento oscuro y por su combinación desagradable de retórica radical y autoritarismo profesoral. Pero hay algo que aprender de ella, particularmente si algunas de sus intuiciones se utilizan en la teoría de la justicia y no en conexión con la libertad, que fue donde puso el énfasis la Escuela de Frankfurt.

[12] Una cuestión importante es la de hasta qué punto una forma universal de justicia puede adoptar un contenido diferente en sociedades diferentes; esta idea es funda-

pensamiento ético que trascienda el relativismo de la distancia. Quizá esto se deba a una característica del mundo moderno. Tenemos diversas concepciones de la justicia social, que tienen diferentes consecuencias políticas; y todas ellas tienen inteligibles raíces en el pasado y en nuestros sentimientos. En tanto que sabemos que no aceptamos sus legitimaciones del pasado, pero que tampoco estamos seguros de cómo entenderlas, estamos dispuestos a considerar las concepciones de la justicia del pasado como la materialización de ideas que todavía tienen algo que decir en el mundo moderno. En este sentido, consideramos que están en confrontación real entre sí y con las ideas modernas.

Retorno ahora a la reflexión misma y a su relación con el conocimiento ético. He dicho que la reflexión podría destruir el conocimiento, porque los conceptos éticos densos que se usaron en un estadio menos reflexivo podrían dejar de usarse a causa de la reflexión, mientras que los pensamientos éticos más generales y abstractos que probablemente ocuparían su lugar no satisfarían las condiciones del conocimiento proposicional. Decir que el conocimiento queda destruido en tal caso no es decir que las creencias particulares que una vez fueron verdaderas dejen ahora de serlo, ni tampoco supone decir que las personas se den cuenta de que nunca habían sabido lo que creían saber. Lo que significa es que esas personas antes tenían creencias de cierto tipo, que en muchos casos constituían conocimiento, pero ahora, debido a que tras la reflexión no pueden ya usar conceptos esenciales para esas creencias, ya no pueden formarse ese tipo de creencias. Un determinado tipo de conocimiento relativo a situaciones concretas, que se usó como guía en su mundo social y contribuyó a formarlo, ya no existe. El conocimiento queda destruido porque la potencialidad para determinado tipo de conocimiento ha sido destruida; además, si piensan sobre sus creencias anteriores, las verán

---

mental para la concepción de la justicia de Michael Walzer en su útil libro *Spheres of Justice* (Nueva York, Basic Books, 1983) [*Las esferas de la justicia*, trad. de H. Rubio (México, FCE, 2004)]. Otra cuestión es la de cómo podemos considerar injusto el pasado al mismo tiempo que sabemos que le debemos casi todo lo que apreciamos. Examino algunas concepciones griegas de la injusticia social y sus relaciones con las nuestras en el capítulo 5 de *Shame and Necessity* (California University Press, 1993) [*Vergüenza y necesidad*, trad. de A. Montes Sánchez (Madrid, A. Machado, 2011)].

ahora como las veía un observador, como conocimiento que ellos no comparten.

Es sabido que la reflexión destruye el conocimiento. Es un lugar común que, en un caso individual, una destreza práctica puede quedar destruida por la reflexión sobre cómo se practica (aunque en circunstancias favorables la reflexión puede igualmente mejorar la práctica). Pero este es un caso muy diferente del que nos ocupa. En primer lugar, se trata solo de la consciencia personal del individuo: los observadores pueden indagar teóricamente la forma en que la persona dotada de una aptitud particular hace uso de ella, y sus conclusiones, si se usan correctamente, pueden mejorar su uso. Una segunda diferencia es la de que la reflexión del individuo tiene que ver con una acción que este indudablemente puede realizar, y aunque puede hacer que no logre realizarla de nuevo, no es necesario que esto sea así. En el caso ético, sin embargo, aunque es verdad que antes de la reflexión las personas podían desenvolverse sin problemas en el mundo social usando esos conceptos, la implicación de la reflexión es que ahora han de hacer otra cosa. A diferencia de las reflexiones inquietantes del ciclista y del funámbulo, la reflexión ética forma parte de la práctica misma que somete a consideración, y la modifica intrínsecamente.

Sócrates pensó sobre estos temas solo desde el punto de vista de la reflexión de un individuo sobre su propia práctica. Pensaba que era imposible que la reflexión destruyera el conocimiento, puesto que no puede haber conocimiento sin reflexión. Creía que si había algo que podía conducir al conocimiento, era la reflexión y que lo importante era el conocimiento (se está mejor con conocimiento que sin él). Si una persona tiene la segunda de esas creencias sin tener la primera, la idea de que la reflexión puede destruir el conocimiento se volverá contra la reflexión y se expresará en el tipo de conservadurismo, o algo peor, que elogia el arraigo, la comprensión tácita y las maneras tradicionales de entender las cosas. Evidentemente, a favor de estas cosas se puede decir más de lo que ha permitido gran parte del pensamiento progresista; en realidad, hay más que decir a favor de estas cosas que a favor de una gran parte del pensamiento progresista. Pero, incluso si se concede valor al conocimiento tradicional, tratar de suprimir la reflexión en su beneficio no puede llevar más que al desastre; más o menos como sucede con quien, considerando que el haber tenido hijos ha trastornado su vida, no puede recuperar su estado anterior más que acabando con ellos.

Pero no habríamos de aceptar la segunda creencia de Sócrates. Para aceptar la paradoja antisocrática, habría que rechazar sus dos presupuestos. El conocimiento ético, aunque exista, no es necesariamente el mejor estado ético. Tenemos que recordar, al respecto, que en el proceso de pérdida del conocimiento ético podemos ganar conocimiento de otros tipos sobre la naturaleza humana, sobre la historia, sobre cómo es el mundo en realidad. Podemos obtener conocimiento sobre, o alrededor de, lo ético. Dentro de la esfera de lo ético, por el mismo proceso, podemos ganar en comprensión.

Este no es solo otro nombre para el conocimiento que habremos perdido. Por encima de todo, ha cambiado su relación con la convicción. Una razón por la que conservadores y tradicionalistas atacan la reflexión es que temen la incertidumbre que parece generar, la situación en la que incluso los mejores carecen de toda convicción. El resultado que temen es algo terrible, y tienen razón en detestar una determinada postura liberal que hace de la incertidumbre misma una virtud y reemplaza la convicción por el disfrute de las satisfacciones —satisfacciones no menos intelectuales— de la indecisión refinada. Pero esos tradicionalistas y esos liberales comparten el error de pensar que, en la vida ética, la convicción tiene que ser conocimiento, que tiene que ser un modo de certeza.

Si no cabe identificar la convicción ética con el conocimiento o con la certeza, ¿entonces en qué consiste? Hay quienes rechazan concebirla como certeza, pero la reemplazan por otra concepción que no es más sensata, sino menos razonable. Pensando que más allá del intelecto solo está la voluntad, creen que el origen de la convicción ética tiene que ser una *decisión*[13]: la decisión de adoptar determinados principios morales

---

[13] La versión más conocida e interesante de esta concepción se encuentra en el existencialismo defendido por Sartre después de la Segunda Guerra Mundial, aunque luego llegó a admitir que era tan ridículo como le había parecido siempre a muchos otros. En una forma menos drástica, la concepción se ha vuelto casi un lugar común en gran parte de la filosofía reciente. John Mackie, por ejemplo (*Ética: la invención de lo bueno y lo malo;* pág. 106 [ed. inglesa; pág. 120, ed. cast.]), pudo decir, sin considerar que era algo muy especial, que «la moral no es algo que deba descubrirse sino algo que debe hacerse: nosotros tenemos que decidir cuáles serán los puntos de vista morales que adoptaremos», pero no está claro si el «nosotros» significa cada uno de nosotros o todos nosotros en conjunto ni, en todo caso, qué tenemos que hacer en un caso o en el otro. Es probable que, en esos pasajes, se le dé una forma psicológica a una doctrina lógica o metafísica, lo cual se presta a confusión.

o de vivir de un modo y no de otro. Esto no puede ser así porque la convicción ética, como cualquier otra forma de *estar convencido,* tiene que contener un aspecto de pasividad, tiene que venir a nosotros, en cierto sentido. Puede parecer que determinadas decisiones son así, pero esto se debe a que son decisiones particularmente persuasivas. No podemos tener una convicción ética y, al mismo tiempo, ser conscientes de que es el producto de una decisión, a menos que la decisión misma parezca inevitable. Pero, en tal caso, esto último es lo que cabría explicar.

Kant creía que la moral requería de la autonomía y que ningún principio moral podría ser propiamente *mío* —o, por decirlo de otro modo, nada que fuera mío podría ser un principio moral—, a no ser que yo lo haya reconocido o adoptado libremente. Pero, como vimos en el capítulo 4, ese era un punto de vista que Kant aplicó también a las conclusiones teóricas, y que sostuvo al nivel de su psicología trascendental. No se ocupaba de las decisiones en tanto que característica psicológica de nuestra experiencia cotidiana. Como veremos en el siguiente capítulo, necesitaba encontrar en la experiencia cotidiana algo que tuviera precisamente el carácter opuesto, algo que acompañara a los juicios de la razón, y esto se presentaba en el modo de la pasividad, como un sentimiento que parecía estar determinado desde fuera, que denominó «el sentimiento de respeto hacia la ley».

Ni el modelo de la decisión ni el de la certidumbre parecen muy útiles frente a la carencia efectiva de convicción ética. Ciertas personas se pronuncian a favor del modelo de la certidumbre afirmando que tenemos necesidad de la convicción ética y que solo el conocimiento puede proporcionárnosla. Pero ignoran el hecho obvio de que ni siquiera una buena dosis de fe en la certeza cognitiva nos suscitará la convicción ética si no podemos concordar en aquello sobre lo que se supone que versa la convicción. Al apreciar la fuerza de esta objeción, los defensores del otro modelo dicen que hemos estado buscando en el sitio equivocado: la ética es una cuestión de decisión y tenemos que afrontar la responsabilidad y asumir la carga de las decisiones. Esta concepción ignora el hecho igualmente obvio de que, si la ética es una cuestión de decisión, y nos embarga la incertidumbre, entonces estamos igualmente inciertos sobre lo que hemos de decidir.

Necesitamos una tercera concepción, para la que la mejor palabra es tal vez la de *confianza.* La confianza es fundamentalmente un fenómeno social. Esto no obsta para que, cuando se da en una sociedad, es porque los individuos la poseen, ni para que no pueda darse en algunos

individuos si está ausente en la sociedad. Cuando esto sucede, no obstante, ocurre de una forma diferente, dado que la ausencia de confirmación y apoyo sociales de las actitudes de un individuo tiene que afectar al modo en el que este mantiene esas actitudes; en primer lugar, llevándolo a percatarse de ellas. La razón de traer a colación esta concepción no es decir que la filosofía, incapaz de decirnos cómo generar convicción, puede mostrarnos cómo generar confianza. Se trata, sobre todo, de que esta concepción muestra más claramente que los otros modelos por qué la filosofía no puede decirnos cómo generarla. Es una cuestión social y psicológica la de qué tipos de instituciones, educación y discurso público contribuyen a fortalecerla. Las primeras cuestiones que hemos de tener presentes sobre la confianza ética se refieren a una explicación social. Esto no significa que la confianza ética no tenga nada que ver con la argumentación racional. Los estados sociales pueden ser afectados, de un modo u otro, por la argumentación racional. Además, si tratamos de generar confianza sin argumentación racional, o suprimiéndola, es bastante probable que fallemos y que, además, sacrifiquemos otros bienes. La confianza es solo un bien entre otros: tiene un precio, y el precio no habría de ser muy alto.

La confianza es un estado social y está vinculada al debate, la teorización y la reflexión; correspondientemente, estas actividades son formas de practicarla que ocupan el espacio social del mismo modo que en el individuo ocupan el espacio psicológico. Se nos induce a olvidar este hecho por medio de una serie de concepciones intelectualistas que pretenden hacernos creer que nuestro objetivo fundamental es el de llegar a responder las preguntas éticas, que para hacerlo hemos de prestar tanta atención como sea posible a las razones poderosas que guardan relación con esas preguntas, y que las exigencias de la práctica que limitan esas actividades se reducen a una limitación práctica. La verdad es que la pregunta esencial es la de cómo vivir y qué hacer, para lo cual son relevantes las consideraciones éticas; y la cantidad de tiempo o de energía humana dedicadas a reflexionar sobre esas consideraciones tiene que depender de lo que, según nuestra perspectiva de la vida ética, consideremos como una vida digna de ser vivida y de lo que podrían hacer las personas para que la vida mereciera la pena. Una pregunta que tenemos que responder es la de cómo las personas, o bastantes personas, pueden llegar a tener una confianza práctica que, incluyendo en particular la necesidad de reflexionar y su presencia generalizada en nuestro mundo, procederá de la fortaleza y no de la debilidad del autoengaño y del dog-

matismo. (La confianza no es lo mismo que el optimismo; podría descansar en lo que Nietzsche llamó pesimismo de la fortaleza).

Este análisis nos sugiere una conclusión sobre el futuro del pensamiento y de la práctica éticos. Hemos de recordar la idea, que consideré al final del capítulo anterior, de que los seres humanos alcancen un fundamento objetivo para la vida ética; o, dicho de una manera que será más útil aquí, la idea de que logren una vida ética que sepan que tiene unos fundamentos objetivos. Esto no puede constituir una aspiración que cuente con una gran probabilidad de hacerse realidad, pero hay algo más que puede aprenderse de la idea de ello. El proceso implicaría una convergencia práctica, sobre una vida ética en común. En el caso de la ciencia, mi concepción de la objetividad incluía la idea de una convergencia *sin coacción:* si no fuera así, no podríamos explicarla como un procedimiento para llegar a la verdad. En el caso de la convergencia práctica, ética, habría que explicar la convergencia en términos de deseos o intereses elementales, y esto también requiere que el procedimiento no sea coactivo. Y se plantea la cuestión de qué contaría como un procedimiento no coactivo.

Algunos procedimientos estarían excluidos, por supuesto. Si nos visitaran los marcianos y dejaran claro que, en caso de que los seres humanos no alcanzaran una alto nivel de acuerdo sobre determinadas formas de vida ética, destruirían el planeta, podría ser que se lograra un acuerdo en un par de generaciones, quizá con la ayuda de determinada tecnología que trajeran los marcianos. Esto implicaría, sin lugar a dudas, deseos e intereses elementales, pero no la idea de fundamentar objetivamente nuestra vida. Ocurriría simplemente que, debido a esas sanciones, esa vida se habría aceptado. Después de que hubiéramos llegado a los estados de cosas exigidos, los alienígenas tendrían que mantenernos conscientes del terror, destruir nuestro poder de reflexión o erigir un mito legitimador poderoso. Si, prescindiendo de todo eso, pudieran hacer que el sistema funcionara satisfactoriamente, tendrían que habernos dado un modo de vida satisfactorio (aunque, quizá, no el único posible), pero lo que lo haría satisfactorio sería el hecho de que nos permitiría vivir de manera estable y reflexiva y no que se nos haya forzado a converger en él.

Si el acuerdo no se debiera a la coacción, tendría que surgir del interior de la vida humana. Y tendría que recibir la influencia, al mis-

mo tiempo, de la indagación teórica. Un proceso de este tipo necesita de instituciones libres, que no solo permitan la indagación libre, sino también la diversidad de formas de vida y cierta variedad ética. No obstante, tenemos que estar alerta, de nuevo, a no caer en el error, en su versión social, de presuponer que la reflexión no ocupa ningún espacio psicológico. Una sociedad entregada a «experimentar en su manera de vivir», según la expresión de Mill, no es una sociedad en la que aumenten las probabilidades de vivir del mejor modo. Es una sociedad de un tipo más que de otro, y descarta diversas formas de vivir; de hecho, entre las rechazadas podrían estar las más dignas de ser vividas. Sin embargo, esto solo implica que la diversidad y la libertad de investigación sean como la confianza, unos bienes que cabe promover entre otros, pero no que no sean bienes. Quienes creen en la objetividad, y ven que la única forma inteligible de objetividad es un fundamento objetivo, tienen razones para aceptar que son bienes importantes.

También tenemos razones para aceptar que se trata de bienes importantes aunque no creamos en la objetividad. Esto se debe a un relativismo de la distancia dirigido al futuro. El caso de quienes vivan en una sociedad futura es el único en el que podemos tener tanto una confrontación puramente nocional con otro conjunto de valores como cierta responsabilidad con respecto a esos valores; una responsabilidad, por lo menos, en la medida en que podemos intentar imprimir nuestros valores en las generaciones futuras o rechazar hacerlo. Para tener confianza en la tarea de asegurarnos de que las generaciones futuras compartan nuestros valores, nos haría falta, me parece, no solo confiar en esos valores —lo que, si podemos conseguirlo, es algo bueno—, sino también estar convencidos de que son objetivos, lo que constituye un error. Sin esta convicción, tenemos razones para abstenernos de influir en el futuro, tanto como las tenemos para no juzgar el pasado. No habríamos de tratar de imprimir determinados valores en la sociedad futura.

Asimismo, tenemos razones para adoptar ciertas medidas positivas. Habríamos de intentar preservar los recursos necesarios para una vida adecuada y, como medio y parte de ello, tratar de transmitir lo que consideramos que es nuestro saber. Si queremos ser coherentes, no podemos excluir la consciencia reflexiva, así como las prácticas de la investigación libre que la sostienen y permiten hacer un uso adecuado de ella. Por supuesto, existen tensiones entre estos dos objetivos, el negativo y el positivo. Tratar de transmitir la libre indagación y la consciencia

reflexiva es transmitir algo en vez de nada, y algo que exige ciertas formas de vida en lugar de otras.

Tenemos razones para intentar transmitir más cosas a nuestros sucesores inmediatos, a nuestros hijos, al menos. Aspirar a su preservación es un signo de que poseemos valores éticos. Pero esto no afecta de un modo muy determinado a lo que mantendrán las generaciones futuras más remotas. Si nuevos desarrollos aumentaran nuestra influencia en sus perspectivas, haríamos bien en no hacer uso de ellos, más allá de dejarles, si podemos, la libre investigación y la reflexión, un legado creado por nuestro conocimiento. Esto será suficiente a modo de legado; no tratar de transmitirles nada más mostrará un respeto apropiado por el relativismo de la distancia.

Capítulo 10

La peculiar institución de la moral

Ya me he referido a la moral como un sistema especial, una variedad particular del pensamiento ético. Tengo que explicar ahora cómo la entiendo yo y por qué estaríamos mejor sin ella.

Lo importante con respecto a la moral es su espíritu, sus objetivos subyacentes y la imagen general de la vida ética que comporta. Para poder discernirlos, necesitaremos examinar cuidadosamente un concepto particular: el de *obligación moral*. No es el mero uso de la noción de obligación lo que hace especial a la moral. Existe una noción ordinaria de obligación, según la cual no es más que una consideración entre otras, que resulta éticamente útil. La moral se distingue por la noción especial de obligación que utiliza, y por el significado que le otorga. Es a esta noción especial a la que denominaré «obligación moral». La moral no es un conjunto determinado de pensamientos éticos, sino que abarca un elenco de perspectivas éticas y está tan incardinada en nosotros que la filosofía moral dedica la mayor parte de su tiempo a discutir las diferencias entre esas perspectivas, en vez de ocuparse de lo que las diferencia de todo lo demás. No todas estas perspectivas son ejemplos igualmente típicos o instructivos del sistema de la moral, aunque tienen en común la idea de obligación moral. Kant es el filósofo que nos ha proporcionado la representación más pura, profunda y rigurosa de la moral. Pero la moral no es una invención de los filósofos. Es la actitud o, de manera menos coherente, parte de la actitud, que tenemos casi todos nosotros.

En el sistema de la moral, la obligación moral se expresa por medio de un tipo especialmente importante de conclusión deliberativa: una conclusión referida a qué hacer, regida por razones morales y relativa a una situación concreta. (Hay también obligaciones de tipo general; volveremos a ellas más adelante). No toda conclusión de una deliberación moral particular, incluso dentro del sistema de la moral, expresa una obligación. Sin ir más lejos, ciertas consideraciones morales afirman meramente que está *permitido* hacer cierta cosa. Estas conclusiones no expresan una obligación, aunque están regidas, en algún sentido, por la idea de obligación: alguien se pregunta si está sujeto a una obligación y decide que no lo está.

Esta descripción está en función del resultado o conclusión de la deliberación moral. Las consideraciones morales que son parte de una deliberación pueden tomar la forma de obligaciones, pero es natural afirmar que no es necesario que lo hagan. Yo podría, por ejemplo, concluir que estaba sujeto a la obligación de actuar de determinada manera, porque lo mejor sería que se diera determinado resultado y yo podría lograr que se produzca actuando de esa manera. Sin embargo, el sistema de la moral ejerce una presión que hace que se tienda a representar como obligación general toda consideración que interviene en la deliberación y genera una obligación concreta. Así pues, si estoy ahora mismo sujeto a la obligación de hacer algo que sería lo mejor es porque tengo una obligación general, quizá entre otras muchas, de hacer lo que es mejor. Veremos más adelante los pormenores de este movimiento.

El hecho de que la obligación moral sea un tipo de conclusión práctica explica algunas de sus características. Una obligación se aplica a alguien con respecto a una acción —es una obligación de hacer algo— y la acción tiene que estar dentro de las posibilidades del agente. «*Deber* implica *poder*» es una fórmula famosa a este respecto. Como enunciado general sobre el *debe* es falsa, pero tiene que ser correcta si se toma como condición de qué puede constituir una obligación particular cuando es lo que se concluye a nivel práctico. Si mi deliberación me conduce a algo que no puedo hacer, tengo que volver a deliberar. La cuestión de qué cosas se considera que están dentro de las posibilidades del agente es notablemente problemática, no solo debido a las numerosas y desconcertantes teorías que sostienen que todo (o todo lo psicológico) está determinado, sino también porque no está claro qué quiere decir que alguien puede actuar, o habría podido actuar, de cierta mane-

ra. Para decir algo de provecho con respecto a estos problemas, sería necesario un debate de gran envergadura que no intentaré desplegar en este libro[1]. No obstante, lo que tengo que decir aquí indicará que, en este y otros respectos, la moral choca con problemas ordinarios de una forma peculiarmente aguda.

Otra característica de las obligaciones morales, en este sentido, es que no pueden entrar en conflicto entre sí, definitiva y realmente o en última instancia. Esto se sigue directamente de la última observación, la de que lo que estoy obligado a hacer tengo que poder hacerlo, si concedemos el principio adicional (que ha recibido el nombre de «principio de aglomeración») de que, si estoy obligado a hacer X y estoy obligado a hacer Y, entonces estoy obligado a hacer X e Y. Este requisito también refleja la forma práctica de esta noción de obligación. En un sentido cotidiano de «obligación», que no está regido por estos requisitos especiales, es evidente que las obligaciones pueden entrar en conflicto. De hecho, una de las ocasiones más comunes para mencionarlas es cuando esto sucede[2].

El filósofo David Ross inventó una terminología, que aún se usa a veces, para clasificar los conflictos de obligaciones, que distinguía entre obligaciones *prima facie* y obligaciones efectivas. Una obligación *prima facie* es una conclusión, que se apoya en consideraciones morales, que pretende ser una obligación efectiva de alguien. Será la conclusión de mi deliberación moral si ninguna otra obligación la supera. Ross trató de explicar (sin mucho éxito) por qué una simple obligación *prima facie* —ante la que finalmente puede prevalecer otra— es algo más que solo una obligación aparente. Cabe pensar que ejerce cierta fuerza en el momento de la decisión, pero no la suficiente, teniendo en cuenta su rivalidad con otras obligaciones, para tomar la decisión. En términos más concretos, esto tiene como efecto que las consideraciones que apoyaban

---

[1] Me ocupo brevemente de algunas de estas cuestiones más adelante en este capítulo. La mayoría de las discusiones sobre el libre albedrío no prestan la debida atención a que la explicación causal puede tener un impacto diferente en distintas partes de nuestro pensamiento sobre la acción y la responsabilidad. Vale la pena recordar que la deliberación requiere solo *poder*, mientras que la censura implica *haber podido*.

[2] He abordado la cuestión del conflicto en diversos artículos, en *Problemas del yo* y en *La fortuna moral*. Es importante notar que, si fuera lógicamente imposible que dos obligaciones efectivas entraran en conflicto, yo no podría encontrarme en una situación en la que entraran en conflicto, aunque sea por mi culpa. ¿Pero en qué situación se supone que me habría metido?

la obligación *prima facie* descartada puedan dar apoyo a otra obligación efectiva. Si, por buenas e imperiosas razones, rompo una promesa puede que adquiera una obligación efectiva de hacer otra cosa, como compensar a la persona a la que he defraudado.

No está del todo claro por qué habría de estar sujeto a esta obligación adicional, dado que es cosa mía, desde este punto de vista, cumplir con mis obligaciones, y eso es lo que habría hecho. No se ha incumplido ninguna obligación efectiva. La consecuencia reconfortante es que no habría de censurarme a mí mismo. Puedo culparme por otra cosa, por meterme en esta situación, pero es erróneo censurarme o reprocharme a mí mismo por no haber realizado la acción desestimada: el autorreproche va a la par con las obligaciones incumplidas, y al fin y al cabo ha resultado que no había ninguna obligación. Se admite que es razonable que me sienta mal por ello, pero el sistema de la moral distingue este sentimiento del remordimiento o el autorreproche, dándole el nombre, por ejemplo, de «lamento», que no sería un sentimiento moral. Esta nueva clasificación es importante y muy característica de lo que sucede cuando se reduce lo ético a lo moral. Decir que los sentimientos que alguien experimenta con respecto a algo que ha hecho involuntariamente, o como el menor de dos males, han de entenderse como formas del lamento, un sentimiento no moral, conlleva que ese individuo se ha de sentir con respeto a esas acciones como se siente con respecto a las cosas que meramente acaecen o a las acciones de otras personas. El pensamiento de que *Yo lo hice* no tiene un significado especial, lo que es significativo es si hice voluntariamente lo que debía haber hecho. Esto desvía el foco de nuestra atención de una dimensión importante de la experiencia ética, que reside sencillamente en la distinción entre lo que uno ha hecho y lo que no ha hecho. Esta puede ser una distinción tan importante como la existente entre lo voluntario y lo no voluntario[3].

La obligación moral es ineludible. Puedo adquirir una obligación voluntaria, como cuando hago una promesa; en ese caso, es usual decir que tiene que hacerse voluntariamente para que sea una promesa, aunque haya una zona de sombra aquí, como cuando prometemos bajo coacción. En otros casos, puedo estar sujeto a una obligación sin haber-

---

[3] Discuto este problema en mi artículo «La suerte moral», que ilustra la idea general de que el sistema de la moral otorga una importancia particularmente grande a la insegura estructura de lo voluntario.

la elegido. Pero, en todo caso, una vez estoy sujeto a una obligación, no puedo escapar a ella, y el hecho de que un agente dado prefiera no estar atado a las reglas de este sistema no le excusará; y censurarlo no será fruto de un malentendido. La censura es la reacción característica del sistema de la moral. El remordimiento, el autorreproche o la culpa, que ya he mencionado, constituyen la reacción característica de primera persona dentro del sistema y, si un agente no experimentara nunca estos sentimientos, no pertenecería al sistema de la moral o no sería un verdadero agente moral en los términos del sistema. El sistema también implica la censura entre las personas y, sin duda, de no existir tal cosa, no se darían esas reacciones de primera persona, pues estas se forman por interiorización. Pero es posible que los agentes particulares que pertenecen al sistema no censuren nunca a nadie, en el sentido de expresar censura y quizá incluso de sentir los sentimientos relevantes. Podrían, por ejemplo, ser escrupulosamente escépticos con respecto a lo que está bajo el poder de otras personas. El hecho de que censurarnos a nosotros mismos o tener remordimientos requiera que nuestra acción haya sido voluntaria es solo una aplicación particular de la regla general de que la censura se dirige a lo voluntario. La ley moral es más exigente que la ley de una república liberal, porque no podemos emigrar —no hay escapatoria—, pero es inequívocamente justa en su concepción de la responsabilidad.

En este sentido, el utilitarismo es un miembro marginal del sistema de la moral. Tiene en su haber una tradición muy arraigada de pensamiento que defiende que la censura y otras reacciones sociales deberían reasignarse con el fin de que sean socialmente útiles y, de la misma forma que esto podría llevar a que se restringieran a lo voluntario, también podría resultar en lo contrario. Esto se sigue lógicamente de la aplicación del criterio utilitarista a todas las acciones, incluidas las acciones sociales con las que se expresa la censura y otras reacciones semejantes. El mismo principio puede extenderse a la censura tácita y a los pensamientos críticos; de hecho, a otro nivel, un utilitarista podría preguntarse si la política más útil no habría de consistir en olvidar que la razón de ser de la censura —para el utilitarismo— es su utilidad. Estas maniobras parecen recibir confirmación cuando se aplican al autorreproche y al sentido de obligación moral. Los utilitaristas suelen ser personas muy coherentes, que trabajan por la humanidad y renuncian a comer carne por el bien de los animales. Piensan que esto es lo que, moralmente, deben hacer, y se sienten culpables si no viven de acuerdo a sus propios criterios. No se preguntan, ni quizá podrían hacerlo, por

qué utilidad tiene pensar y sentir así. Es debido a estas motivaciones, y no solo a sus rasgos lógicos, por lo que el utilitarismo es, en la mayoría de sus versiones, un tipo de moral, aunque sea uno marginal.

El sentimiento de que la obligación moral es ineludible, que lo que estoy obligado a hacer es lo que *tengo que* hacer, es la manifestación en primera persona de la concepción, ya mencionada, de que la obligación moral se nos aplica aunque no queramos. En tercera persona, esta concepción supone que el juicio moral y la censura pueden aplicarse a todos, incluso aunque, en el caso límite, quieran vivir fuera de ese sistema. Desde la perspectiva de la moral no hay un lugar fuera del sistema, por lo menos no para el agente responsable. En la terminología de Kant, podemos unir esos dos aspectos y decir que la obligación moral es *categórica.*

Volveré a la cuestión de las personas que están fuera del sistema más adelante. Ahora es preciso concretar aún más lo que es una obligación moral para alguien que está dentro del sistema. Es difícil estar de acuerdo con que aquel curso de acción para cuya adopción, en cierta ocasión, tenemos más razones morales tenga necesariamente que constituir una obligación moral. Hay acciones (así como políticas, actitudes, etc.) que son o algo más o algo menos que las obligaciones. Pueden ser acciones heroicas o sublimes, que van más allá de lo que es obligatorio o exigible. O pueden ser acciones que, desde el punto de vista ético, sería agradable o valioso o una buena idea realizar sin que se esté obligado a realizarlas. La cuestión es obvia en términos de las reacciones de las personas. Podemos admirar mucho a algunas personas, o tenerlas en gran estima, por acciones cuya omisión no suscitarían nuestra censura. ¿Cómo aborda el sistema de la moral las consideraciones que aparentemente no dan origen a obligaciones?

Un modo en el que la versión central, deontológica, de la moral trata estas consideraciones es intentando convertir en obligaciones a tantas como sea posible. (Es un motivo particular de la empresa reduccionista tratar de asimilar todas las consideraciones éticas a un solo tipo). Hay algunos ejemplos instructivos de esto en la obra de Ross, cuya terminología de las obligaciones *prima facie* ya he mencionado. Ross enumera diversos tipos de las que considera obligaciones generales o, como él dice, deberes generales[4]. El primer tipo comprende lo que

---

[4] W. D. Ross, *The Right and the Good* (Oxford, Clarendon Press, 1930), págs. 21 y ss. [*Lo correcto y lo bueno,* trad. de L. Rodríguez (Salamanca, Sígueme, 1994)].

todos llamamos una obligación, a saber, guardar las promesas y, como extensión natural, decir la verdad. La segunda clase comprende los «deberes de la gratitud»: hacer el bien a quienes nos han prestado un servicio. Pero no está del todo claro que esto sean *deberes,* a no ser que el benefactor (como puede implicar la palabra «servicio») haya adquirido un derecho a esperar algo a cambio —en cuyo caso se seguirá de una promesa implícita, y la obligación pertenecerá al primer tipo—. Las buenas obras que no he pedido pueden resultarme opresivas, pero no he de confundir esa opresión con una obligación[5].

Lo que Ross trata de hacer entrar en el molde de la obligación es una idea ética, sin duda, diferente: que es signo del buen carácter querer devolver los favores. Esta característica no es igual a una disposición a hacer lo que se está moralmente obligado a hacer. La tercera clase de Ross, que denomina los «deberes de la justicia», encubre, de nuevo, una idea ética distinta. Lo que dice sobre esto es sorprendente: «[estos deberes] se basan en el hecho o la posibilidad de una distribución de placer o felicidad (o de los medios para alcanzarla) que no esté en consonancia con los méritos de las personas implicadas; en tales casos surge el deber de invertir o evitar semejante distribución»*.

Hay cosas tales como los deberes o las obligaciones de la justicia, pero esta incitación a la insurrección contra la economía capitalista (o contra cualquier otra economía, por cierto) es difícil que constituya una explicación correcta de lo que son. Las exigencias de la justicia conciernen, en primer lugar, a *lo que debe suceder.* La manera en que una determinada exigencia de la justicia se relaciona con lo que una persona dada tiene razones para hacer o, más específicamente, tiene la obligación de hacer, depende de la relación que mantenga esta persona con esta exigencia. En política, una de las principales cuestiones es la de hasta qué punto la acción personal se aparta de lo deseable —la *medida de la utopía,* como podría denominarse—, y esta es también una de las principales cuestiones éticas.

Es un error de la moral el tratar de convertirlo todo en obligaciones. Pero las razones de este error son profundas. Cabe recordar aquí

---

[5] Esto es así aun cuando las buenas obras formen parte de una práctica general a la que los demás esperan que me una. La cuestión la presenta admirablemente Robert Nozick en *Anarchy, State, and Utopia* (Nueva York, Basic Books, 1974) [*Anarquía, Estado y utopía,* trad. de R. Tamayo (México, FCE, 1991)], cap. 5.

* Trad. cast., pág. 36. *[N. del T.].*

que lo que denominamos *de ordinario* una obligación no tiene por qué imponerse necesariamente en caso de un conflicto entre consideraciones morales. Supongamos que alguien tiene una obligación ordinaria: visitar a un amigo, digamos (un ejemplo de manual), porque se lo ha prometido. A esta persona se le presenta la oportunidad, en el momento y lugar del conflicto, de promover significativamente alguna causa importante. (Para hacer más realista el ejemplo, sería necesario abundar en los detalles, pero, como sucede a menudo en la filosofía moral, si damos más detalles, el ejemplo puede empezar a desvanecerse. Está la cuestión de la actitud de su amigo hacia la causa y también hacia el apoyo que da esa persona a la causa. Si es favorable a ambas cosas, o solo a la segunda, y si liberaría a la persona de la promesa en caso de poder ponerse en contacto, solo el moralista más rigorista hallaría una dificultad en esto. Si el amigo no estuviera dispuesto a librarle de la promesa podría preguntarse qué tipo de amigo tiene... Pero no es difícil que cada cual interprete esto de tal forma que encuentre un ejemplo claro que ilustre este argumento). La persona puede concluir razonablemente que ha de aprovechar la oportunidad de promover la causa[6]. Pero las obligaciones tienen un rigor moral, lo que implica que incumplirlas acarrea ser censurado. Lo único con lo que puede contarse para cancelar esto, dentro de la economía de la moral, es que la acción rival represente otra obligación más estricta. La moral fomenta la idea de que *solo una obligación puede prevalecer sobre una obligación*[7].

---

[6] Es un ejemplo de un conflicto entre una obligación y una consideración que a primera vista no es una obligación. También puede representar muy fácilmente el conflicto entre lo privado y lo público. Para diversas disquisiciones al respecto, y particularmente sobre el papel de las consideraciones utilitaristas en la vida pública, véanse los artículos incluidos en Stuart Hampshire (ed.), *Public and Private Morality* (Nueva York, Cambridge University Press, 1978).

[7] Sin duda, la moral fomenta esta idea, en casos de este tipo, pero no siempre insiste en ella, por lo menos en la forma de que una obligación mía solo puede quedar relegada por otra obligación mía. Si un interés vital mío tiene que ser sacrificado para cumplir una promesa, y la promesa es relativamente poco importante, incluso el moralista más severo admitirá que tengo derecho a incumplir la promesa, sin que sea necesario que tenga la obligación de hacerlo (debo esta consideración a Gilbert Harman). Esto es correcto, pero, a no ser que la promesa sea muy trivial, sospecho que el moralista estricto solamente lo admitirá si los intereses en juego son de hecho vitales. Esto sugiere una interpretación según la cual mi obligación sería superada por una obligación, pero no por una de las mías. Al insistir en que solo cuentan los intereses vitales, es probable que el moralista, cuando dice que tengo derecho a salvaguardar mis intereses, no quiera

Con todo, ¿cómo puede esta acción de esa persona haber sido una obligación, a no ser que proceda de alguna obligación más general? No es fácil decir en qué consiste esta obligación general. La persona de la que hablamos no está sujeta a una obligación indefinida de perseguir esa causa ni de hacer todo lo que le sea posible para defender las causas que haya adoptado. Nos queda la idea poco convincente de que esa persona está sujeta a la obligación de contribuir a una causa importante en las ocasiones especialmente propicias para ello. La exigencia interna al sistema de la moral de encontrar una obligación general que respalde a la obligación particular —lo que podríamos denominar el principio de *una obligación, otra obligación*— ejerce una presión que produce un resultado más claro en los casos cotidianos en los que una emergencia centra en una circunstancia particular una consideración ética general, como la obligación de tratar de socorrer a alguien en peligro. No estoy obligado a socorrer a toda persona que corra un riesgo ni a buscar a personas en peligro para prestarles asistencia. Confrontados[8] con alguien que corre peligro, muchos se sienten en la obligación de tratar de prestar ayuda —aunque sin correr un peligro excesivo para ellos, etc.; hay diversas condiciones razonables—. En este caso, a diferencia del precedente, la obligación subyacente parece hecha a medida. La exigencia inmediata que se me aplica («En esta emergencia, tengo la obligación de ayudar») se entiende que deriva de «Tenemos esta obligación general: prestar socorro en caso de emergencia». Si añadimos la idea de que muchas consideraciones morales, quizá todas, podrían prevalecer sobre cierta obligación en una ocasión dada, hallamos que muchas de esas consideraciones, tal vez todas ellas, están relacionadas con algunas consideraciones generales, incluso si no son las obligaciones simples e indeterminadas que proponía el reduccionismo de Ross.

---

decir solo que puedo hacer eso, sino que tengo lo que se ha llamado un «derecho de reivindicar» su ejecución, por el cual los demás están sujetos a la obligación de no impedirme hacerlo. Así, mi obligación inicial será cancelada por la obligación del beneficiario de la promesa a renunciar a su derecho de exigir que se cumpla.

[8] Qué cuenta como confrontación es una cuestión real y muy práctica para los médicos en particular. Abordo brevemente el asunto más adelante, al dar una explicación de la inmediatez que no necesita del principio de *una obligación, otra obligación*. Es evidente que se trata de un tipo de obligación cada vez más ignorada en las ciudades modernas, en la medida en que no será bien recibida ni siquiera por las personas que, al sentirse culpables, hacen mutis por el foro.

Una vez que se ha iniciado el camino hacia las obligaciones más generales, comienzan los problemas —no solo filosóficos, sino también de conciencia— para dar cabida a las acciones moralmente indiferentes. He mencionado ya la posible conclusión moral de que alguien *puede* adoptar cierto curso particular de acción. Eso significa que no hay nada más que esté obligado a hacer. Pero, si hemos aceptado las obligaciones generales e indeterminadas de promover diversos objetivos morales, como las últimas observaciones nos animan a hacer, será necesario promoverlos y no dedicarnos a la ociosidad, y puedo tener motivos (no estoy diciendo que los tenga) para pensar que puedo estar mejor ocupado con esto que dedicándome a hacer cosas que no tengo la obligación de hacer y, si puedo estarlo, entonces debería: tengo la obligación de no desperdiciar el tiempo haciendo cosas que no tengo obligación de hacer. En esta fase, sin duda, solo una obligación puede prevalecer sobre una obligación y, para hacer lo que quiero hacer, me hará falta uno de esos instrumentos fraudulentos, un deber hacia mí mismo. Si se deja que la obligación estructure el pensamiento ético, habrá diversos modos naturales en los que podrá llegar a dominar totalmente la vida.

Para comprender la estructura intimidatoria que la moral ha hecho de la idea de obligación hace falta explicar qué son las obligaciones cuando acertadamente se las considera un tipo de consideración ética entre otras. Esta explicación nos ayudará a desprendernos de la noción de obligación moral específica del sistema de la moral y, en último término, a escapar por completo de este sistema.

En primer lugar, nos hace falta la noción de *importancia.* Como es bastante obvio, la importancia de las cosas varía de una persona a otra (lo que no significa necesariamente que esas cosas sean importantes para los intereses de esas personas). Esto comporta una noción relativa de importancia, que también podríamos caracterizar diciendo que alguien *encuentra* importante determinada cosa. Más allá de esta idea meramente relativa, tenemos otra noción, por la que algo es, sin más, importante (importante *überhaupt,* como dirían otros, o importante *y punto).* No está nada claro qué es que algo sea, sin más, importante. No significa que sea importante para el universo: en ese sentido, nada es importante. Ni significa que, como cuestión de hecho, sea algo que la mayoría de seres humanos encuentran importante, ni que sea algo que las personas deben encontrar importante. Dudo que pueda haber

una explicación incontestable de esta idea; las explicaciones que las personas dan al respecto se ven necesariamente afectadas por lo que encuentran importante.

No importa, para nuestros propósitos, que no comprendamos bien esta noción. Solo necesito que se reconozcan tres cosas; una es que esa noción existe. Otra es que, si una cosa es importante para alguien, en el sentido relativo, esto no necesariamente implica que piense que es, sin más, importante. Puede ser de la mayor importancia para Henry completar su colección numismática con un determinado sello, pero el mismo Henry puede entender que esto no es, sin más, importante. Aquí reside una idea esencial: las personas han de encontrar importantes muchas cosas que son, sin más, importantes, así como muchas cosas que no lo son, y han de poder indicar la diferencia entre ellas.

El tercer elemento es que la cuestión de la importancia, y sobre todo la cuestión de qué es, sin más, importante, tiene que distinguirse de la cuestión de la *prioridad deliberativa*. Una consideración tiene gran prioridad deliberativa para nosotros si le damos más peso que a las demás consideraciones en nuestras deliberaciones. (Esto contiene dos ideas: que, cuando interviene en nuestras deliberaciones, contrarresta en peso a la mayoría de las otras consideraciones, y también que interviene en nuestras deliberaciones. Hay ciertas razones para ocuparnos de la segunda idea por separado, y aludiré a una de estas razones después, pero en general resulta más simple tratarlas conjuntamente).

La importancia tiene algunas conexiones con la prioridad deliberativa, pero no son directas. Hay muchas cosas importantes sobre las que uno no puede hacer mucho, y muchísimas en relación con las cuales una persona dada no puede hacer nada. Por otro lado, puede que no sea de la incumbencia de esa persona hacer nada al respecto; hay una división deliberativa del trabajo. Mis deliberaciones no están conectadas de una manera sencilla ni siquiera con lo que es importante para mí. Si encuentro que algo es importante, esto afectará a mi vida y, por tanto, a mis deliberaciones, de una manera u otra, pero esos efectos no tienen por qué hallarse directamente en el contenido de mis deliberaciones.

Una consideración puede tener una gran prioridad deliberativa para una persona particular, para un grupo de personas o para todo el mundo. De esta manera, la prioridad se torna relativa a las personas. Pero no habría de relativizarse en otros respectos: no habría de estar marcada por su objeto, de modo que determinadas cosas tuvieran una prioridad deliberativa moral o prudencial. Esto sería incurrir en

una confusión. Podría decirse que las consideraciones morales tienen una gran prioridad desde el punto de vista moral. Si esto es así, lo que significará es que quien que está dentro del sistema de la moral da una gran prioridad a esas consideraciones. Pero no define un tipo de prioridad. Una cuestión esencial de la prioridad deliberativa es que puede relacionar consideraciones de diferentes tipos[9]. Lo mismo puede decirse de la importancia. En un sentido, hay diferentes tipos de importancia, y es natural decir que algunas cosas son moralmente importantes, otras son estéticamente importantes, etc. Pero al final, en un caso particular o en general, se planteará la cuestión de si un tipo de importancia es más importante que otro.

Quienes están dentro del sistema de la moral suelen pensar que la moral es importante. Además, la moral tiene que ver, por definición, con la conducta personal, por lo que aquí la importancia es probable que tenga que ver con la deliberación. Pero qué tiene que ver con qué depende crucialmente del modo en el que uno entiende la moral y su importancia. Para los utilitaristas, lo importante es que haya tanto bienestar como sea posible. La conexión con la deliberación es una cuestión subsiguiente, que está enteramente abierta. Vimos al examinar el utilitarismo indirecto que la cuestión está abierta en relación con qué consideraciones morales han de formar parte de las deliberaciones del agente utilitarista, e incluso en relación con si se requiere que alguna consideración moral forme parte de ellas. Cierto tipo de utilitarismo ha supuesto que los mejores resultados se darían si las personas no pensaran en términos morales en absoluto y simplemente (por ejemplo) actuaran de forma egoísta. Con menos fe en la mano invisible, otros dan prioridad a las consideraciones morales, y algunos de ellos, como hemos visto, adoptan una línea muy cautelosa. Pero para cualquier utilitarista siempre ha de haber una cuestión empírica: ¿cuáles son las implicaciones para la deliberación de que el bienestar sea importante? En este aspecto, sin embargo, hay muchos utilitaristas que pertenecen al sistema de la moral en primer lugar, y solo son utilitaristas en segundo lugar.

En el otro extremo, los modelos más puramente kantianos sitúan la importancia de la moral en la importancia de la motivación moral misma. Lo importante es que las personas otorguen la mayor prioridad

---

[9] Esto está relacionado con el examen de las cuestiones deliberativas del capítulo 1.

deliberativa a las consideraciones morales. Esta concepción fue correcta e implacablemente atacada por Hegel sobre la base de que no daba ningún contenido al pensamiento moral e implicaba una ambigüedad sobre la forma de mejorar el mundo. El contenido de la motivación moral era la obligación de hacer determinadas cosas, frente a la mera inclinación; implicaba necesariamente que los individuos no estaban espontáneamente inclinados a hacer esas cosas; su importancia suprema residía en que era mejor así.

Ninguna de estas concepciones es satisfactoria, y ningún compromiso bastará para mejorarlas. La vida ética misma es importante, pero puede apreciarse que hay otras cosas que también son importantes. Contiene motivaciones que sirven, de hecho, a esos otros fines, pero que al mismo tiempo pueden verse desde dentro de esa vida como parte de lo que la hace digna de ser vivida. Se mire como se mire, las motivaciones éticas serán importantes, y esto tiene consecuencias para cómo hemos de deliberar. Una de las consecuencias es que algunos tipos de consideración ética tendrán gran prioridad deliberativa. Esto solo es una manera en la que las motivaciones éticas pueden afectar a las deliberaciones de las personas. También pueden afectar a su estilo y a sus circunstancias, entre otras cosas[10].

Hay un tipo de consideración ética que conecta directamente la importancia y la prioridad deliberativa: la obligación. Se funda en la cuestión básica de en qué personas se ha de poder confiar. En la medida de lo posible, las personas tienen que confiar en que no serán asesinadas o utilizadas como un medio y en tener un espacio, unos objetos y unas relaciones con otras personas que puedan considerar propios. También sirve a sus intereses si, en alguna medida al menos, pueden contar con que no se les mienta. Un modo en el que se puede servir a esos fines, tal vez el único, es mediante algún tipo de vida ética; y, sin duda, si hay vida ética, esta tiene que servir a esos fines y estos tienen que lograse en ella. *Una* manera en la que la vida ética sirve a esos fines es alentando determinadas motivaciones, y *una* forma de conseguir esto es inculcar una disposición a otorgar gran prioridad deliberativa a las consideraciones relevantes —en los casos más serios, una prioridad virtualmente absoluta, de modo que se prioricen determinados cursos

---

[10] Es igualmente pertinente recordar algo que se dijo en el capítulo primero: las consideraciones deliberativas que acompañan a una motivación ética dada, tal como una virtud, pueden no tener ningún vínculo con ella.

de acción, mientras se descartan otros desde el principio—. Una manera efectiva para descartar acciones, que suele ser la mejor, es no pensar nunca en ellas. No nos sentimos cómodos con el hombre que, en el transcurso de una discusión sobre cómo tratar con rivales en política o en los negocios, dice que «Por supuesto, podríamos hacer que los matasen, pero hemos de comenzar por descartar esta eventualidad». Nunca debería habérsele ocurrido descartar esta eventualidad. Es característico de la moral que tienda a pasar por alto la posibilidad de que ciertas inquietudes se expresan mejor de este modo, con el silencio deliberativo.

Las consideraciones a las que se les otorga prioridad deliberativa para asegurar la fiabilidad constituyen obligaciones, a las cuales corresponden derechos, que poseen las personas que se benefician de las obligaciones. Ciertas obligaciones se singularizan por la importancia básica y permanente de los intereses a los que sirven. Todas estas obligaciones tienen una fuerza negativa que se aplica a lo que no hemos de hacer. Otras obligaciones, positivas, son las obligaciones que son producto de la inmediatez. Aquí, la inmediatez impone una gran prioridad deliberativa, como en el caso del rescate que vimos antes. La inmediatez hace que un reconocimiento ético general de los intereses vitales de las personas se centre en una prioridad deliberativa; es lo que para *mí* resulta inmediato lo que genera *mi* obligación, una obligación que no puedo ignorar sin ser censurable. Esta manera de comprender las obligaciones debidas a la inmediatez tiene dos consecuencias. En primer lugar, después de todo, no tenemos por qué decir que la obligación procede de una obligación más general. El punto clave de las obligaciones negativas no está en que sean generales, sino en que proporcionan un modelo permanente e invariable de prioridades deliberativas. En el caso de las obligaciones positivas, sin embargo, la disposición subyacente es una preocupación general que no se expresa siempre en términos de prioridad deliberativa, y lo que produce la obligación es, precisamente, la inmediatez. No tenemos por qué aceptar el principio de *una obligación, otra obligación.*

Lo que es más importante, esta manera de comprender esas obligaciones tiene consecuencias éticas. Algunos moralistas dicen que si tenemos en cuenta la inmediatez o la proximidad física es por falta de imaginación o de racionalidad; somos irracionales si no reconocemos que quienes pasan hambre en otros lugares tienen derecho a esperar lo mismo de nosotros que quienes pasan hambre aquí. Estos moralistas se equivocan, al menos al intentar basar sin más el reto que plantean en la

estructura de las obligaciones. Ni que decir tiene que esto no responde al reto mismo. Sin duda, hemos de prestar más atención a los sufrimientos de las personas de otros lugares. Pero una comprensión correcta de las obligaciones nos mostrará cómo hemos de pensar sobre esta problemática. No se trata de erradicar la categoría de la inmediatez, sino que tenemos que considerar qué ha de contar como inmediato para nosotros, en el mundo moderno, y qué lugar ocupan en nuestras vidas este tipo de preocupaciones cuando no son obligaciones.

Las obligaciones consideradas hasta el momento implican (negativamente) lo que es fundamentalmente importante y (positivamente) lo que es importante e inmediato. Ambas cosas se basan, en último término, en la idea de que toda persona tiene una vida que vivir. Las personas necesitan ayuda, pero no en todo momento (a menos que sean muy jóvenes, muy mayores o tengan una discapacidad grave). No tienen permanentemente la necesidad de no ser asesinados, o de no sufrir atracos, ni de padecer interferencias arbitrarias en sus vidas. Uno de los puntos fuertes del contractualismo es haber visto que las obligaciones positivas y negativas derivan de esos intereses básicos[11].

Las obligaciones con las que estamos más familiarizados, las debidas a promesas, difieren de las precedentes en que lo que estoy obligado a hacer puede, en sí mismo, no tener ninguna importancia. Pero, justo por eso, son ejemplos de la misma conexión entre la obligación y la fiabilidad. La promesa es una institución que funciona de tal forma que proporciona una fiabilidad transferible, ofreciendo una fórmula que conferirá una gran prioridad deliberativa a lo que, en otro caso, podría no tenerla. Es por esto que resulta extraño que alguien prometa no matar a otra persona. Si no le confiere una gran prioridad a esto, ¿por qué habríamos de confiar en que la promesa sí se la conferirá? (Hay respuestas para esta cuestión, en casos especiales, y reflexionar sobre cuáles podrían ser ayudará a mostrar cómo funciona el sistema).

La función de la obligación es asegurar la fiabilidad, esto es, un estado de cosas en el que las personas pueden esperar razonablemente que los demás se comporten de una manera y no de otra. Es solo una forma ética de actuar entre otras, que trata de producir la expectativa *de que* mediante una expectativa *de*. Tales obligaciones exigen muy a menudo

---

[11] La referencia al contractualismo pone de relieve el carácter, en algún sentido, individualista de lo que se ha expuesto. Para otras consideraciones al respecto, véase el Epílogo.

la mayor prioridad deliberativa y se presentan también a sí mismas como importantes; en el caso de las promesas, no solo por su contenido, sino porque son promesas. Sin embargo, podemos también advertir que no siempre necesitan tener la mayor prioridad, incluso en agentes con buena disposición ética. Reflexionando sobre el hecho de que un objetivo es especialmente importante, y de que la acción presente está relacionada con él de manera importante, un agente puede concluir razonablemente que la obligación puede incumplirse en esta ocasión, como se apuntó antes, y esta conclusión puede ser aceptable[12], en el sentido de que él mismo puede explicar dentro de una estructura de consideraciones éticas por qué decidió lo que decidió. Pero no es necesario que designe a este curso de acción como una obligación diferente y más rigurosa. Una obligación es un tipo especial de consideración, que guarda una relación general con la importancia y la inmediatez. El caso que estamos considerando es simplemente uno en el que una consideración es tan importante que contrarresta esta obligación en esta ocasión[13], y resulta más clarificador decir sencillamente esto. Hemos de rechazar la otra máxima de la moral: que solo una obligación puede prevalecer sobre una obligación.

Una conclusión deliberativa que incorpora una consideración que tiene tanto la mayor prioridad deliberativa como la mayor importancia (por lo menos para el agente) puede adoptar una forma especial y convertirse en la conclusión no solo de que uno ha de hacer algo en concreto, sino de que *tiene que* hacerlo y de que no puede hacer otra cosa. Podemos llamar a esto una conclusión debida a la necesidad práctica. A veces, por supuesto, en una conclusión práctica «tener que» posee solo un valor relativo y significa meramente que cierto curso de acción es necesario para alcanzar un fin que no es, en absoluto, cosa de «tener que». «Tengo que marcharme ya» puede bien completarse con «...si quiero ir al cine», donde no se sugiere que tenga que ir necesariamente al cine: me limito a querer ir al cine. Aquí no nos interesa este «tener que», sino un «tener que» incondicional que *llega hasta el fondo.*

---

[12] Es un error suponer que tiene que ser igualmente aceptable para todo el mundo. Ciertas personas pueden tener más derecho que otras a quejarse.

[13] ¿Es *tipo* de ocasión? Sí. Pero los hechos singularizadores, tales como que sea la segunda vez (para ella, este año), pueden sin duda ser pertinentes.

Una cuestión interesante es la de en qué medida una conclusión concerniente a lo que tenemos que hacer, o a lo que no podemos hacer, difiere de una conclusión que solo afirma que tenemos más razones para hacer algo; en concreto, ¿cómo es que la primera conclusión puede ser más fuerte, como parece que es el caso? (¿Cómo la deliberación puede conducirnos a una conclusión favorable a un curso de acción que es más firme que las razones que tenemos a su favor?) No trataré de examinar esta cuestión aquí[14]. Lo que nos importa ahora es que la necesidad práctica no es en ningún caso específica de la ética. Alguien puede concluir que tiene que hacer algo incondicionalmente, por razones de prudencia, de autoprotección, de interés estético o artístico o de pura autoafirmación. En algunos de estos casos (la autodefensa básica, por ejemplo), una perspectiva ética puede autorizar la conclusión. En otros casos, la desaprobará. La cuestión fundamental es que una conclusión que resulta de la necesidad práctica es el mismo tipo de conclusión esté o no fundada en razones éticas.

La necesidad práctica, y la experiencia de alcanzar una conclusión con esa fuerza, es un elemento que se ha vinculado a la idea de obligación moral (lo que puede ayudar a explicar la sensación compartida por muchos de que la obligación moral es a la vez muy especial y resulta muy familiar). Pero la necesidad práctica, aun cuando se fundamente en razones éticas, no señala necesariamente una obligación. El curso de acción que el agente «tiene que» tomar puede no estar asociado a las expectativas de los demás, ni con la censura cuando no se realiza. Las acciones éticamente excepcionales o heroicas que mencioné antes, al ser más que obligaciones, no son obligatorias, y normalmente no se nos puede pedir que las llevemos a cabo, ni se nos puede reprochar que no las realicemos. Pero el agente que realiza una acción semejante puede tener la sensación de que tiene que hacerlo, que no tiene alternativa, y al mismo tiempo reconocer que no podría exigírselo a los demás. Si bien puede pensar que es una exigencia para él, no para los demás, porque él es diferente a los demás, normalmente la diferencia radicará en el hecho de que es justamente él quien tiene esa convicción. Sus sentimientos, y las expectativas sobre qué sentimientos experimentará si no actúa de ese modo, pueden aseme-

---

[14] He hecho una propuesta al respecto en «La necesidad práctica», artículo incluido en mi libro *La fortuna moral*.

jarse mucho a los asociados a las obligaciones (más de lo que la moral admite)[15].

Ya me he referido a la descripción que hace Kant de la moral como categórica. Al afirmar que el principio fundamental de la moral era el imperativo categórico, Kant no estaba interesado en una distinción puramente lógica entre las formas de lo que literalmente son imperativos, sino en el reconocimiento de un *tengo que* incondicional llevado hasta sus últimas consecuencias, pero interpretó esta necesidad práctica incondicional como peculiar de la moral. Kant pensaba que era incondicional en el sentido de que no depende en absoluto del deseo: un curso de acción que se nos presenta con este tipo de necesidad es tal que tenemos razones para adoptarlo *independientemente de lo que podamos querer,* y solo las razones morales podrían trascender el deseo de este modo. No obstante, tal y como la he presentado, no es necesario que la necesidad práctica sea independiente del deseo en un sentido tan fuerte. He distinguido un «tener que» incondicional de uno *condicionado por un deseo eventual del agente,* pero una conclusión relativa a la necesidad práctica podría ser la expresión de un deseo, si este deseo no fuera uno más entre los que tiene el agente, sino uno cuya satisfacción fuera esencial para el agente. La diferencia entre esta concepción de la necesidad práctica y la de Kant no es, por supuesto, solo una cuestión de definición o de análisis lógico. La idea que tiene Kant de la necesidad práctica es básicamente esta otra, con la que estamos más familiarizados, aunque dándole una interpretación particularmente radical: las únicas conclusiones prácticas necesarias son las absolutamente incondicionadas con respecto a todo deseo. Para Kant, podía haber una conclusión práctica que fuera radicalmente incondicionada de este modo, a causa de su caracterización del yo racional como libre de la causalidad, y porque hay razones para la acción que dependen mera-

---

[15] ¿Cómo de semejantes? Esto afecta a una cuestión importante que no puedo examinar aquí: la distinción entre la culpa y la vergüenza. Me ocupo en profundidad de ello en *Vergüenza y necesidad,* esp. cap. 4 y nota final 1. Esta distinción existe y es relevante para la ética, pero es mucho menos compleja de lo que se suele pensar. Sobre todo, es un error suponer que la culpa puede identificarse con una reacción madura y autónoma que tiene un lugar en la experiencia ética, mientras que la vergüenza es una reacción más primitiva que no lo tiene. La moral tiende a equivocarse con respecto a su relación con la vergüenza. Para observaciones sugerentes sobre esta distinción, véase Herbert Morris, «Guilt and Shame», en *On Guilt and Innocence* (Berkeley, University of California Press, 1976).

mente de la agencia racional y no de algo (como un deseo) que el agente podría no haber tenido[16].

Kant también describe las conclusiones de la necesidad práctica, concebida como específica de la moral, como el reconocimiento de las exigencias de la ley moral y, cuando habla de esto en términos psicológicos, se refiere a un sentimiento especial, un «sentimiento de respeto hacia la ley». Es poco probable que los moralistas modernos usen estas palabras, pero no les resulta difícil reconocer lo que Kant describe. (Algunos de ellos todavía quieren invocar una concepción de la ley moral. Otros, que son reluctantes a hacerlo, usan ideas que implícitamente la invocan). Kant no pensaba que el sentido irresistible de necesidad moral, concebido como un sentimiento, era en sí el que proporcionaba la razón de la acción moral. Como sentimiento, no era más que un sentimiento, sin más poder racional que cualquier otro elemento meramente psicológico. La razón no residía en la naturaleza de ese sentimiento, sino en lo que representaba, esto es, la verdad de que la universalidad moral es un requisito de la misma razón práctica.

Para Kant, esa verdad significaba que la moral tiene un fundamento objetivo —como vimos en el capítulo 4— y la experiencia de la exigencia moral representaba este fundamento. Sin embargo, hay que decir también que Kant da una interpretación errónea muy significativa de esa experiencia. La experiencia consiste en verse *confrontado* con alguna cosa, con una ley que forma parte del mundo en el que uno vive[17]. Pero el poder de la ley moral, según Kant, no se encuentra, ni es

---

[16] Esto está conectado con las concepciones divergentes del yo de Kant y de sus críticos hegelianos; véase el capítulo 1, nota 6. Es importante distinguir aquí dos ideas distintas: otras personas, y yo mismo, podemos tener una idea «externa» de diferentes ideales y proyectos a los que yo podría haber tenido, por ejemplo, si hubiera sido educado de forma diferente; hay pocas razones a favor, y muchas en contra, para sostener que, si yo hubiera sido educado de forma diferente, no habría sido yo. Este es el dominio de la necesidad metafísica. Pero existe otro dominio, el de la necesidad práctica, que se interesa por qué líneas de acción y qué proyectos son posibles para mí, dados los ideales y el carácter que de hecho tengo. Este es el nivel al que tenemos que resistirnos a la idea kantiana de que el auténtico sujeto ético es aquel para el que nada es necesario, excepto la agencia misma. Esto está también estrechamente relacionado con el problema de los intereses reales, discutido en el capítulo 3.

[17] El modelo de la ley moral permite explicar por qué el sistema ha de tener las dificultades que tiene con aquellos actos éticos que, como dije antes, son más o menos que las obligaciones. No es sorprendente que algo que se concibe como una ley haya de presentar solo las tres categorías de lo exigido, lo prohibido y lo permitido. Los intentos

concebible que pueda encontrarse en nada externo a uno mismo. Su poder radica en su fundamento objetivo, y ninguna experiencia podría representar adecuadamente este tipo de objetividad. La objetividad procede del hecho de que las exigencias de la razón práctica serán satisfechas solo llevando una vida en la que las consideraciones morales jueguen un papel fundamental y característico; y ese papel lo realizarán solo si, a diferencia de otras motivaciones, presentan la forma de una exigencia objetiva. Pero, ¿qué significa que una consideración se presente como una exigencia objetiva? No puede consistir en que se presente como tal en relación con ese mismo argumento. Ha de tener otra forma psicológica y, en esa medida, la forma será equívoca.

Los presupuestos kantianos, sin embargo, permiten comprender cómo y por qué tal experiencia está destinada a ser confusa, y esto contribuye a que sea un objeto estable de reflexión. Si Kant tiene razón, yo puedo llegar a comprender lo que es el «sentimiento de respeto hacia la ley» sin perder mi respeto por ese sentimiento o por la ley moral. Otra idea contribuye a esta estabilidad, a saber, la idea de que hay un sentido en el que la experiencia representa correctamente la ley como exterior a mí, porque se encuentra igualmente implantada en otras personas. La ley moral es la ley de la república nocional de los agentes morales. Se trata de una república nocional, pero los agentes son reales y, dado que se trata de una ley que cada agente se impone racionalmente a sí mismo, es una ley real.

Una vez hemos dejado de creer en el fundamento en el que Kant creía, o en algo semejante, no podemos ya seguir interpretando esta experiencia de ese modo. Es la conclusión de la necesidad práctica, ni más ni menos, y parece proceder «de afuera» del modo en el que esas conclusiones de la necesidad práctica siempre parecen proceder de afuera —de lo más profundo de nosotros—. Dado que están en juego consideraciones éticas, las conclusiones del agente no estarán normalmente aisladas ni carecerán de fundamento, pues forman parte de una vida ética que, en gran medida, compartimos con otras personas. Desde este punto de vista, el propio sistema de la moral, que insiste en lo «puramente moral» y en los sentimientos personales de culpa y de

---

de Kant de hacer frente a algunos problemas de esos otros motivos éticos dentro de su marco del deber se plasman en sus versiones (pues cambió su concepción con el tiempo) de la distinción tradicional entre deberes perfectos e imperfectos. Sobre esto, véase M. J. Gregor, *Laws of Freedom* (Nueva York, Barnes & Noble, 1963), caps. 7-11.

autorreproche, oculta la medida en que la vida ética se sitúa fuera del individuo.

Cuando sabemos lo que es el reconocimiento de la obligación, si continuamos haciendo de la obligación el centro especial de la experiencia ética, construiremos la vida ética sobre una ilusión. Incluso desde la perspectiva de Kant esta experiencia supone una mala interpretación, que es necesaria y aceptable, pues es consecuencia de la trasposición de la objetividad del ámbito trascendental al psicológico. Pero si esta experiencia solo es especial en el modo psicológico, el resultado es peor que una mala interpretación: no hay nada (o nada especial) que represente.

La construcción kantiana también explica cómo la ley puede aplicarse incondicionalmente a todas las personas, incluso a las que tratan de vivir al margen. Quienes no aceptan esta construcción, pero aceptan el sistema de la moral, necesitan dar cuenta de cómo quienes rechazan la obligación moral están sujetos a ella. Hace falta que expliquen, en primer lugar, cómo puede existir una *ley* moral[18]. Para que una ley se aplique a alguien no basta con una relación semántica; no se trata solo de que la descripción contenida en la ley se aplique a la persona. La ley del Estado se aplica a una persona porque esta pertenece a un Estado que puede ejercer el poder. La ley de Dios se aplica porque Dios la aplica. La ley moral de Kant se aplica porque, en tanto que ser racional, tengo una razón para aplicármela a mí mismo. Para que se aplique la ley moral solo se necesita que nosotros la apliquemos.

Cuando decimos que alguien debía haber actuado de la manera exigida o deseable y que no lo ha hecho, a veces decimos que *había una razón* para que actuara de ese modo —lo había prometido, por ejemplo, o bien que lo que hizo violaba los derechos de otra persona—. Aunque podemos decirlo así, esto no parece guardar una conexión inequívoca

---

[18] La cuestión del imperativo categórico y su relación con las razones para actuar ha sido abordada por Philippa Foot en distintos artículos, compilados en *Virtudes y vicios*. Debo mucho a esos artículos, aunque mis conclusiones son diferentes. El *debe* moral fue uno de los diversos blancos atacados por G. E. M. Anscombe en su vigoroso artículo «Modern Moral Philosophy», reimpreso en *Ethics, Religion and Politics*, vol. 3 de sus *Collected Papers* (Minneapolis, University of Minnesota Press, 1981) [en castellano en M. Platts (ed.), *Conceptos éticos fundamentales* (México, UNAM-IIF, 2006), págs. 27-53; trad. del artículo de L. E. Manríquez].

con la idea de que *tenía una razón* para actuar así. Tal vez no tenía ninguna razón. Al incumplir la obligación, no estaba comportándose necesariamente de forma irracional o irrazonable, sino de forma incorrecta. No podemos dar por sentado que él tenía una razón para comportarse bien, más allá de que nosotros tengamos diversas razones para desear que se hubiera comportado bien. ¿Cómo lo tratamos? Reconocemos de hecho —muy chapuceramente en la ley, menos en la práctica informal— que hay muchas maneras distintas en las que las personas pueden dejar de ser lo que éticamente nos gustaría que fuesen. En un extremo está la incapacidad deliberativa general; en el otro, el devoto sincero y competente de otro credo. Si bien hay personas con diferentes debilidades o vicios, personas que son maliciosas, egoístas, brutales, desconsideradas, autoindulgentes, perezosas, codiciosas... Todas estas personas pueden formar parte de nuestro mundo ético. Ningún mundo ético ha estado nunca libre de quienes tienen esos vicios (aunque su clasificación dependerá de la cultura en cuestión), y cualquier vida individual está pautada por algunos de ellos. También hay diversas reacciones negativas ante ellos, desde el odio y el horror, en los casos más extremos, hasta la indignación, el lamento, la corrección, la censura... Cuando no estamos inmersos en las circunstancias formales de la ley del Estado se plantea la cuestión adicional de quién reacciona: no todo el mundo puede o ha de expresar cualquier tipo de queja. Otra consecuencia de la ficción de la ley moral es que esta verdad no se nos haga patente. Es como si todo miembro de la república nocional estuviera facultado para arrestar a un ciudadano.

Entre todo esto, hay un número bastante significativo de desviaciones particulares que manejamos mediante los mecanismos de la censura cotidiana. Estas desviaciones incluyen muchos incumplimientos de obligaciones, pero no todos: algunos de los procederes más monstruosos, que sobrepasan la censura ordinaria, entrañan violaciones de derechos humanos fundamentales. Y, por otro lado, la censura no se limita a las obligaciones incumplidas; en particular, en la educación de los hijos se censuran acciones que solo manifiestan disposiciones imperfectas. Pero la censura siempre tiende a compartir el carácter particularizado, práctico, de la obligación moral en el sentido técnico. Su reacción negativa se centra escrupulosamente en una acción u omisión, que es lo que se censura. Además, aunque su funcionamiento contiene muchas anomalías inevitables, la aspiración de la censura es intervenir solo cuando el resultado no deseado es el producto de una acción voluntaria en una ocasión particular.

Esta institución, en contraste con otros tipos de reacciones éticamente negativas u hostiles hacia los actos de las personas (es esencial recordar que hay otras), parece tener que ver especialmente con la idea de que el agente tenía una razón para actuar de otro modo. Como ya he dicho, esto no suele ser así[19]. La institución de la censura se entiende mejor si la consideramos una ficción, en virtud de la cual tratamos al agente como alguien para quien las consideraciones éticas relevantes son razones. El «habrías debido» de la censura puede considerarse un desarrollo para díscolos del «habrías debido» con el que podemos aconsejar a aquellos cuyos fines compartimos. Esta ficción tiene diversas funciones, una de las cuales es la de que, si tratamos al agente como alguien que otorga un peso a las razones éticas, este proceder puede contribuir a convertirlo en ese tipo de persona.

Esta estratagema es especialmente eficaz para mediar entre dos posibilidades en las relaciones entre las personas. Una es la de las prácticas deliberativas compartidas en las que, en gran medida, las personas tienen las mismas disposiciones y colaboran para alcanzar conclusiones prácticas. La otra situación posible es que un grupo aplique la fuerza o amenace con coaccionar al otro. La ficción subyacente al sistema de censura contribuye, en el mejor de los casos, a tender un puente entre estas posibilidades, mediante un proceso de continua adscripción a una comunidad deliberativa. En el peor de los casos, puede tener efectos nefastos, tales como alentar a las personas a interpretar incorrectamente sus propios temores y resentimientos —sentimientos que pueden ser pertinentes— como si fuesen la voz de la Ley.

La ficción de la comunidad deliberativa es uno de los logros positivos del sistema de la moral. Como ocurre con otras ficciones, nos podemos preguntar si su funcionamiento podría soportar una comprensión clara de su forma de proceder. Esto forma parte de la cuestión mucho más general de qué necesita y puede ser reestructurado a la luz

---

[19] Por supuesto, gran parte de la cuestión depende de qué cuente como tener una razón. No creo que puedan existir razones para la acción absolutamente «externas», un tipo de razón que no conecta con ninguna de las motivaciones que ya tiene el agente (como subrayé, Kant tampoco lo creía). Se puede distinguir entre, por ejemplo, limitarse a dirigir la atención del agente a una razón que ya tiene, por un lado, y persuadirle de que actúe de cierta manera, por el otro. Pero lo importante es que existe un abanico de posibilidades, y que esas distinciones son menos claras de lo que requiere el sistema de la moral y otras concepciones racionalistas. Véase «Razones internas y externas», en mi libro *La fortuna moral*.

de una comprensión reflexiva y desmitificada de nuestras prácticas éticas. Es cierto que la práctica de la censura y, en general, el estilo de las reacciones éticas negativas de las personas hacia los demás, cambiarán. El sistema de la moral, desde mi punto de vista, no puede seguir contribuyendo a que esto acaezca de forma satisfactoria. Una razón de ello estriba en que la moral se centra demasiado en la esfera de lo voluntario.

En la medida en que la institución de la censura funciona coherentemente, lo hace porque sus aspiraciones son menos ambiciosas de lo que le gustaría a la moral. Cuando nos planteamos si alguien actuó voluntariamente nos estamos preguntando, a grandes rasgos, si realmente actuó, si sabía lo que hacía o si intentaba tal o cual aspecto de lo que sucedió. Esta práctica considera al agente y a su carácter conjuntamente, y no se cuestiona su libertad para haber elegido tener otro carácter. El sistema de la censura se centra, la mayor parte del tiempo, en las condiciones del acto particular, y puede hacerlo porque no opera por cuenta propia. Está circundado por otras prácticas de estímulo y disuasión, de aceptación y rechazo, que inciden en el deseo y el carácter para adaptarlos a las exigencias y a las posibilidades de la vida ética.

La moral ignora este entorno y solo tiene en cuenta el juicio focalizado y particularizado. El sistema ejerce presión exigiendo una voluntariedad total, que sortee el carácter y la determinación psicológica o social y asigne la censura y la responsabilidad sobre la base, en última instancia, equitativa de la propia contribución del agente, ni más ni menos. Es una ilusión suponer que esta exigencia puede ser satisfecha (a diferencia de los menos ambiciosos requisitos de voluntariedad que conciben el carácter como algo que, en gran medida, nos viene dado). Este hecho es conocido por casi todo el mundo, y es difícil augurar un gran futuro para cualquier sistema que lo niegue. Pero, en tanto que persista la moral, es peligroso admitirlo, puesto que el propio sistema nos deja, como única alternativa a la censura racional, con formas de persuasión que rechaza distinguir, en espíritu, de la fuerza y la coacción.

En verdad, casi toda vida humana valiosa se ubica entre los extremos que la moral nos impone. Esta última acentúa pronunciadamente una serie de oposiciones: entre la fuerza y la razón, la persuasión y la convicción racional, la vergüenza y la culpa, el desagrado y la desaprobación, el mero rechazo y la censura. La actitud que la lleva a acentuar todas estas oposiciones puede etiquetarse como su *pureza*. La pureza de la moral, su insistencia en abstraer la consciencia moral de otros tipos de reacción emocional o influencia social, oculta no solo los medios con

los que trata a los miembros desviados de su comunidad, sino también las virtudes de esos medios. No es sorprendente que tenga que ocultarlos, puesto que las virtudes no pueden considerarse tales más que desde fuera del sistema, desde un punto de vista que pueda conferirle valor, mientras que el sistema de la moral se cierra sobre sí mismo y tiene que considerar un error inaceptable aplicar al sistema otros valores que no sean los de la moral misma.

La propia pureza de la moral representa un valor; expresa un ideal, formulado por Kant, una vez más, y lo hace de la forma más rotunda y una de las más motivadoras: el ideal de que la existencia humana pueda ser en último término justa. La mayoría de las ventajas y de las características que son admiradas están distribuidas de maneras que, si no son injustas, en todo caso no son justas, y algunas personas son sencillamente más afortunadas que otras. El ideal de la moral es un valor, el valor moral, que trasciende la suerte. Por lo tanto, tiene que estar más allá de cualquier determinación empírica. Tiene que consistir solo en el hecho de intentarlo y no en el de conseguirlo, puesto que conseguir algo depende en parte de la suerte, pero tiene que ser un tipo de intento que se sitúe más allá de la medida en la que la capacidad misma de intentar puede ser cuestión de suerte. El valor tiene que ser, además, supremo. No servirá de nada que el valor moral sea meramente un premio de consolación reservado a quien no es feliz o no tiene talento, ni buen humor, o no es querido. Ha de ser lo que importa en último término.

Es semejante, en cierta medida, a una concepción religiosa, pero no se parece a ninguna religión real, ni en particular a la ortodoxia cristiana. La doctrina de la gracia significó para el cristianismo que no había una vía calculable que llevara del esfuerzo moral a la salvación; la salvación se encontraba más allá del mérito, y los esfuerzos de los hombres, incluso sus esfuerzos morales, no eran la medida del amor de Dios[20]. Además, cuando el cristianismo afirmaba que lo últimamente importante era la salvación, se entendía que esto implicaba una diferencia que cualquier persona reconocería como diferencia, como *la* diferencia. Pero el punto de vista desde el que el valor moral puro obtiene su valor es, una vez más, solo el que proporciona la propia moral. Esta solo puede esperar trascender la suerte volviéndose sobre sí misma.

---

[20] Es por eso por lo que dije en el capítulo 4 que la concepción de Kant se parecía a la herejía pelagiana, según la cual la salvación se obtiene solo por el mérito.

Sin lugar a dudas, los ideales de la moral —a diferencia del marxismo vulgar, que los vería solo como una ideología de lo intemporal— han desempeñado un papel importante en la consecución de cierta justicia efectiva en el mundo, así como en la presión sobre el poder y la mejora de las circunstancias sociales para conseguir compensar la mala suerte de manera concreta. Pero la idea de un valor que escapa a la suerte es una ilusión, y los objetivos políticos no pueden continuar extrayendo sus convicciones de esta idea. Una vez más, las otras concepciones de la moral no pueden ayudarnos. Solo pueden alentar la idea, que siempre tiene sus partidarios interesados, de que, cuando se esfuman esas ilusiones, no puede haber ninguna idea coherente de la justicia social, sino tan solo la eficacia, o el poder, o la pura suerte.

La moral está entretejida de muchos errores filosóficos. Concibe erróneamente las obligaciones, al no apreciar que son solo uno de los tipos de consideraciones éticas. Malinterpreta la necesidad práctica, al entenderla como algo peculiar a lo ético y no comprende la necesidad práctica ética, al considerarla peculiar de las obligaciones. Pero, sobre todo, la moral hace creer a las personas que, sin sus obligaciones específicas, solo hay lugar para las inclinaciones; sin su voluntariedad absoluta solo puede haber fuerza; sin su justicia últimamente pura no hay justicia en absoluto. Sus errores filosóficos son solo las expresiones más abstractas de una concepción equivocada de la vida que está profundamente enraizada y es aún poderosa.

Epílogo

El argumento de este libro oscila constantemente entre dos tensiones. Con respecto a las cuestiones filosóficas y a los intereses centrales de la filosofía, se produce una tensión entre lo antiguo y lo moderno. En la vida real, la tensión se da entre la reflexión y la práctica. He defendido que, en ciertos aspectos básicos, el pensamiento filosófico del mundo antiguo era más satisfactorio y se planteó preguntas mucho más fructíferas que la mayor parte de la filosofía moral moderna. Aunque tenía sus limitaciones, tales como el deseo de reducir la vulnerabilidad de la vida con respecto a la suerte, poseía, en general, un talante menos obsesivo que la filosofía moderna, estaba menos determinado a imponer la racionalidad por medio de una teoría reductiva. Sin embargo, las esperanzas que algunos de sus representantes pusieron en poder dar respuesta a estas preguntas se han desvanecido, y el mundo al que se aplica ahora el pensamiento ético es irreversiblemente distinto, no solo del mundo antiguo, sino también de cualquier mundo en el que los seres humanos hayan tratado de vivir y de emplear conceptos éticos.

Los recursos de la mayor parte de la filosofía moral moderna no se ajustan bien al mundo moderno. He tratado de mostrar que esto se debe, en parte, a que está mediatizada en exceso, y de manera demasiado inconsciente, por ese mundo, apelando irreflexivamente a concepciones administrativas de la racionalidad. En alguna de sus modalidades, sobre todo en sus formas más kantianas, no se implica lo suficiente; está embargada por el sueño de una comunidad de la razón que —como denunció Hegel antes que nadie— se aparta demasiado de la realidad

social e histórica y de cualquier sentido concreto de una vida ética particular; más alejada de esas cosas, en ciertos aspectos, que la religión a la que vino a reemplazar. Estas diversas versiones de la filosofía moral comparten una imagen errónea de la relación existente entre la reflexión y la práctica, una imagen de las teorías por medio de las cuales elaboran inútilmente sus diferencias mutuas.

No es una paradoja que, en estas nuevas circunstancias, las filosofías muy antiguas puedan tener más que ofrecernos que las relativamente nuevas, y podría contarse un relato histórico que mostrara por qué es así, que incluiría el advenimiento y la desaparición del cristianismo (lo cual ayuda a explicar por qué el mundo antiguo está más cerca de nosotros de lo que pueda parecer), así como los errores de la Ilustración (que hacen inútiles sus filosofías características). Algunos, quizá la mayoría, de quienes han llegado a esta conclusión —especialmente Nietzsche, en la medida en que cae bajo esta descripción—, han considerado que aniquilaba los valores de la Ilustración; o, si no, como muchos hegelianos, han interpretado esos valores de manera conservadora. No creo que estemos forzados a seguirlos. El respeto por la libertad y la justicia social y la crítica de las instituciones opresivas y engañosas quizá no sean más fáciles de lograr de lo que lo fueron en el pasado, y bien puede que sean más difíciles, pero no tenemos por qué suponer que nos faltan ideas para darles un fundamento. No habríamos de ceder a la pretensión que tiene la teoría ética de que solo ella puede proporcionar el armazón intelectual que esas ideas necesitan.

En este libro me he atenido a lo que es, más que a lo que podría ser, y las esperanzas que he expresado son, por ahora, solo esperanzas. Descansan sobre presupuestos que algunos considerarán optimistas y que pueden sintetizarse en mi creencia en estas tres cosas: en la verdad, en la veracidad y en el significado de la vida individual. Terminaré diciendo algo sobre cada una de ellas.

He defendido que al menos las ciencias naturales son capaces de llegar a la verdad objetiva. El error de quienes niegan esto se expresa característicamente por medio de una retórica del consuelo poco apropiada. Pretenden que aquellos que creen que la ciencia puede decirnos cómo es realmente el mundo se aferran supersticiosamente a ella, con una fe desesperada que hace de la ciencia el único objeto sólido que les queda. Pero también se puede decir que se busca el

consuelo en la dirección opuesta y que el escepticismo contra la ciencia sirve, como hizo en el siglo XVII, para reconfortar a quienes sus propias pretensiones de conocimiento o la racionalidad de sus prácticas les parecen débiles en comparación. La idea de que la ciencia moderna es como habría de ser el conocimiento absoluto puede resultar inquietante, y puede suponer un alivio representarse la ciencia como un conjunto más de rituales humanos o como si no fuera más que otro conjunto de textos.

Estas tesis sobre lo que aporta consuelo o lo elimina pueden defenderse en la dirección que uno desee, y ninguna de ellas tiene el más mínimo interés. Lo más importante, y que puede tener algo que ver con el consuelo o con el optimismo, es saber en qué medida las nociones relativas a la verdad objetiva pueden aplicarse también a la comprensión de la sociedad. Como mencioné en el capítulo 8, no hace falta tratar de ponerse a la altura de las ciencias naturales proporcionando una concepción absoluta del mundo; pero, para que se cumplan nuestras aspiraciones más elevadas de autocomprensión, lo que necesitamos es un conocimiento social reflexivo que incluya la historia y pueda hacerse acreedor de un asentimiento libre de prejuicios. Este conocimiento nos hará falta para realizar el tipo de crítica que nos proporcione una comprensión ética de las instituciones mediante la explicación de cómo funcionan y, en particular, de cómo generan credibilidad. Vale la pena repetir aquí algo que ha surgido en diversas ocasiones al examinar la teoría ética positiva: esta teoría no ha de suponer que puede prescindir de la comprensión de la sociedad ni que puede desentenderse de esos problemas mediante la adopción de una postura puramente moralista. La teoría ética necesita esta comprensión para responder a las preguntas inevitables sobre sí misma: su relación con la vida social, sus conexiones sociales o psicológicas con la práctica y las formas en que es de esperar que su supuesta autoridad racional se transforme en poder.

Consiguientemente, la expectativa de veracidad consiste, en esencia, en que el pensamiento ético aguante la reflexión y que sus instituciones y prácticas se vuelvan transparentes. He tratado de decir por qué el pensamiento ético no tiene la más mínima posibilidad de ser todo lo que aparenta ser. Incluso si el pensamiento ético se fundamentara en determinadas concepciones del bienestar, las consecuencias que se siguen solo justifican *una disposición a aceptar* determinadas afirmaciones éticas, en vez de demostrar, directamente, la verdad de esas afirmaciones; pero naturalmente los que las aceptaran verían las cosas de otro

modo. Además, es improbable que podamos encontrar un fundamento al pensamiento ético.

A pesar de que el pensamiento ético nunca se presentará del todo tal y como es, y de que no puede poner de manifiesto que descansa en disposiciones humanas, los obstáculos para la reflexión que esto supone serán mayores en unas circunstancias más que en otras. Un factor que marcará la diferencia es la medida en la que la vida ética puede descansar aún en lo que he denominado conceptos éticos densos. Estos conceptos son susceptibles de ser desestabilizados por la reflexión, pero, si la resisten, cualquier práctica que los utilice será más estable ante las reflexiones generales, estructurales, sobre la verdad de los juicios éticos que una práctica que los ignore. Los juicios que dependen de estos conceptos pueden ser manifiestamente verdaderos y, correspondientemente, quienes poseen esos conceptos sancionan lo que supone asentir a estos juicios.

Espero haber dejado claro que el ideal de transparencia y el deseo de que nuestra práctica ética sea capaz de resistir a la reflexión no exige que esto se haga de manera totalmente explícita ni que la reflexión aspire a explicarlo todo de golpe. Estas exigencias se basan en una mala comprensión de la racionalidad, tanto personal como política. Tenemos que rechazar todo modelo de pensamiento práctico personal que pretenda que todos mis proyectos, propósitos y necesidades se hayan de convertir, discursiva y simultáneamente, en consideraciones *para* mí. Tengo que deliberar *a partir de* lo que yo soy. La veracidad exige confiar también en esto, y no en la tendencia obsesiva y nefasta a eliminarlo.

¿Cómo cabe combinar la veracidad hacia un yo o una sociedad existentes con la reflexión, la autocomprensión y la crítica? Esta es una cuestión que la filosofía no puede responder, sino que encuentra su respuesta en la vida reflexiva. La respuesta tiene que descubrirse, o establecerse, como el resultado de un proceso, personal y social, que esencialmente no puede formular la respuesta por adelantado, excepto de un modo inespecífico. La filosofía puede desempeñar un papel en el proceso, como lo desempeña la identificación de la cuestión, pero no puede sustituirlo. Es por esto por lo que se equivocan los teóricos éticos cuando preguntan, como suelen hacer, por «qué alternativas» se proponen a sus formulaciones. Lo que quieren decir es qué formulación se propone como alternativa a sus formulaciones, o bien de la respuesta o bien de ciertos procesos heurísticos determinados que nos proporcionarían la respuesta; y no hay ninguna. Podría resultar que hubiera una

respuesta a la pregunta real, y esto constituiría una alternativa a las formulaciones de los teóricos éticos, pero no sería una respuesta elaborada de la manera en que su exigencia necesita una respuesta, como tarea filosófica. Suponer que si rechazamos sus formulaciones, no nos queda *nada* es adoptar una concepción insólita de lo que cuenta como algo en la vida social y personal.

La imagen que ofrezco como trasfondo de esas expectativas requiere, por último, que haya individuos con disposiciones de carácter y con una vida propia que poder vivir. (Como dije en el capítulo 1, esto no nos compromete con la perspectiva biográfica que suscribían los griegos, para los que una vida natural completa es lo éticamente fundamental). En cierto sentido, la primacía de las disposiciones individuales y personales es una verdad necesaria —por lo menos, hasta el advenimiento de los drásticos cambios tecnológicos tales como la clonación, los bancos de cerebros, etc.—. Este es el sentido en el que incluso las descripciones estructuralistas radicales de la sociedad, sean las que sean, suponen que existen individuos que adquieren determinadas disposiciones y metas que expresan en la acción. Si los estructuralistas tienen razón, esas disposiciones están mucho más determinadas por factores sociales tales como la clase, tienen un contenido más uniforme y son menos comprensibles por parte de los individuos de lo que se ha supuesto tradicionalmente; pero esas tesis no pueden negar la existencia y el papel causal de las disposiciones. Ningún conjunto de estructuras sociales puede conducir a los jóvenes a la violencia en los partidos de fútbol si esta no está presente, de manera más o menos confusa u oscura, en los deseos y los hábitos de vida de esos jóvenes. En este sentido, la vida social o ética tiene que existir en las disposiciones de la gente. Es el contenido de esas disposiciones, su inteligibilidad y su grado de particularidad, lo que difiere entre las distintas sociedades y lo que se debate en las diferentes interpretaciones de la sociedad moderna.

Con todo, un individualismo menos formal que ese es seguramente necesario para que sea posible un pensamiento distintivamente ético, que contraste con la planificación social o el ritualismo comunitario; y, con respecto a las esperanzas que estoy expresando aquí, resultará obvio que está en juego un individualismo más sustancial. Así ocurre en otras partes del libro; por ejemplo, en mi caracterización de las obligaciones como unas consideraciones éticas entre otras, así como en ciertas de las afirmaciones que he hecho sobre la necesidad práctica. Mi tercera creencia optimista ha consistido en la posibilidad permanente de una

vida individual significativa, una vida que no rechaza la sociedad y que comparte profundamente sus percepciones con otras personas, pero que es lo suficientemente diferente en sus opacidades y sus desórdenes, y también en sus intenciones razonadas, para hacer que sea la vida de *alguien*. La filosofía puede contribuir a hacer posible una sociedad en la que la mayoría de las personas puedan llevar esa vida, incluso si todavía necesita aprender la mejor manera de hacerlo. La filosofía puede incluso ayudar a ciertas personas a llevar a la práctica esa vida, aunque no a toda persona reflexiva, como suponía Sócrates, ni puede hacerlo de abajo arriba.

Comentario al texto
A. W. Moore

Introducción[1]

Existe un consenso bastante generalizado de que *La ética y los límites de la filosofía* es la obra más sobresaliente de Bernard Williams. Sin ninguna duda, constituye el *locus classicus* de su filosofía moral. Cuando Williams empezó a escribir en esta área, a principios de los años sesenta del pasado siglo, la disciplina se había enzarzado durante bastante tiempo en debates abstractos de segundo orden concernientes al lenguaje moral; sobre si, por ejemplo, un acto de condena moral, tal como decir a alguien «es reprobable que actúe así», constituía una aseveración genuina. Williams estaba decidido a restablecer el contacto con los intereses reales que alientan nuestras experiencias éticas cotidianas. *La ética y los límites de la filosofía* es, en muchos aspectos, la culminación de una cruzada maravillosamente eficaz para lograr exactamente eso. La obra

[1] Este comentario es una versión abreviada y ligeramente adaptada de «Bernard Williams: Ethics and the Limits of Philosohy», en John Shand (ed.), *The Twentieth Century: Quine and After*, volumen 5 de *Central Works of Philosophy* (Chesham, Acumen, 2006). Estoy muy agradecido a John Shand y a Acumen por darme el permiso para reproducir el material de ese ensayo. También estoy muy agradecido a John Shand y a Anita Avramides por sus comentarios a una versión preliminar. Lo que se incluye aquí es, más o menos, un resumen del texto, pensado para ayudar a los lectores a orientarse y como compendio, junto con algunas sugerencias de lecturas complementarias.

muestra admirablemente lo que la filosofía moral puede conseguir. Hay una profunda ironía, por lo tanto, en el hecho de que uno de los temas principales del libro, anunciado en la segunda parte de su título, sea *lo poco* que la filosofía moral es capaz de conseguir. En concreto, la filosofía moral no puede darnos justo lo que se podría haber esperado que nos diera: una *teoría* que guíe el razonamiento ético. Lo que sí puede hacer es contribuir a la autocomprensión por parte de aquellas personas cuyo razonamiento ético ya sigue ciertas directrices; esto es, puede contribuir a proporcionar una crítica de la experiencia ética vivida. Y ello, como insinúa la primera parte del título, es precisamente lo que Williams quiere hacer en estas páginas.

Williams utiliza tanto el Prefacio como el Epílogo para reflexionar sobre diversas facetas de lo que espera conseguir. Hace hincapié en ambos casos en la aparente paradoja de que, aunque, como cree, «las exigencias del mundo moderno al pensamiento ético no tienen precedentes» (pág. 19), también piensa que el pensamiento ético moderno está peor equipado para responder a esas exigencias que el pensamiento ético antiguo. Esta es una de las dos apreciaciones principales que realiza en el Prefacio. La segunda cuestión concierne al estilo, a saber, si este libro puede considerarse una obra de filosofía «analítica». No es que Williams se preocupe mucho por esto; está más interesado, como señala más adelante, en si su libro tiene la virtud que más aprecia la filosofía analítica: la claridad. Y sí tiene un tipo de claridad. Aunque no el tipo que hace más fácil la lectura. Williams nunca se explaya en lo obvio, y casi nunca explicita lo que considera implícito en lo que ya ha dicho. Su prosa es, por ello, extremadamente densa. El lector tiene una ingente tarea por delante. La claridad de su escritura reside en su contenido: es la claridad de la comprensión la que proporcionará al final una recompensa al esfuerzo del lector.

En el Epílogo, Williams escribe que las esperanzas expresadas en el libro «pueden sintetizarse en mi creencia en estas tres cosas: en la verdad, en la veracidad y en el significado de la vida individual» (pág. 242). Y explica, a continuación, lo que quiere decir con ello. En primer lugar, espera que el tipo de autocomprensión que pretende promover pueda responder plenamente a la verdad; en particular, a la verdad relativa a nuestras coordenadas sociales e históricas. En segundo lugar, espera que nuestra experiencia ética pueda soportar esa autocomprensión, incluso si esa autocomprensión nos dice que nuestra experiencia ética no es lo que parece. En tercer lugar, espera que, en caso de que nuestra

experiencia ética resista esa autocomprensión, esto deje libertad a los individuos para hallar, y para dar, sentido a sus vidas. Pese al escepticismo de Williams sobre el poder de la filosofía, el presente libro es una contribución a la realización de estas tres esperanzas.

Capítulo 1. La pregunta de Sócrates

Willliams comienza con una pregunta que denomina «la pregunta de Sócrates», dado que es este quien la plantea en la *República* de Platón (352d). Como dice el propio Sócrates, la pregunta no es trivial. Es nada menos que la pregunta de *cómo uno ha de vivir*.

Desde el principio, Williams deja claro lo poco que habríamos de esperar de la filosofía con respecto a esta pregunta: a buen seguro, no hemos de esperar una respuesta. Pero la filosofía puede ayudarnos a comprender la pregunta. Una gran parte del capítulo 1 está dedicada a examinar la pregunta de Sócrates y, concretamente, a desentrañar sus presupuestos. Al parecer de Williams, la pregunta presupone bastante poco, por lo que es el mejor punto de partida para la filosofía moral. Pero esto no significa que no presuponga nada, insiste Williams. Una cosa que sí presupone es, por supuesto, que los asuntos relativos a cómo vivir pueden abordarse adecuadamente con un grado elevado de generalidad —si no es que se supone que existe algo así como «la vida apropiada [...] para los seres humanos en cuanto tales» (pág. 44)—.

Algo que la pregunta de Sócrates *no* presupone, sin embargo, es lo que Williams denomina «la moral», entendida como un estilo particular de pensamiento ético al que retorna en el último capítulo y que concibe como un rasgo ubicuo y pernicioso del mundo moderno. Mientras que «la ética» es solo otra manera de llamar a la filosofía moral, y está interesada, por lo tanto, en todos los modos de abordar la pregunta de Sócrates; «la moral» —según la útil contraposición que propone Williams entre ambos términos— es una aproximación particular a la pregunta de Sócrates que utiliza unas determinadas herramientas conceptuales muy características[2]. Dos de las más básicas de

[2] Aunque la contraposición es útil, la terminología no lo es tanto, y el lector tiene que ser consciente de que muchos usos habituales de la palabra «moral» y sus derivados, que Williams sigue, tienen más que ver con lo que este denomina «ética» que con lo que denomina «moral». El ejemplo más evidente es la expresión «filosofía moral».

estas herramientas son la idea de acto puramente voluntario y la de obligación moral. La moral interpreta la pregunta de Sócrates como una pregunta sobre qué actos puramente voluntarios tenemos la obligación moral de realizar y cuáles tenemos la obligación moral de abstenernos de realizar, y trata la obligación moral como una exigencia ineludible que eclipsa cualquier otra consideración.

Williams pone en cuestión ambas ideas. Piensa que la idea de un acto que es «puramente» voluntario, así como las demás ideas del juego de herramientas conceptual de la moral con las que se relaciona —la responsabilidad, la culpa, la censura y otras semejantes—, son «una ilusión» (pág. 240), y le molesta la impertinencia y la arrogancia que encuentra en la idea de obligación moral. Hay todo *tipo* de consideraciones, sostiene Williams, que pueden aducirse en relación con la pregunta de Sócrates a parte de las relativas a la obligación; entre ellas, consideraciones éticas de otras clases, como las referidas al bienestar general y a la virtud, y también consideraciones no éticas, como las estéticas y las concernientes al interés propio.

Tampoco cabe pensar que ni las consideraciones éticas ni las no éticas pueden, en último término, reducirse a un único tipo básico. El tema dominante de este capítulo es que cualquier respuesta realista a la pregunta de Sócrates tiene que reflejar la multifacética complejidad de la vida misma.

Capítulo 2. El punto arquimédico

He mencionado la contraposición que hace Williams entre las consideraciones «éticas» y las «no éticas». Williams se abstiene premeditadamente de dar una definición explícita de este contraste; un contraste que considera tanto intuitivo como vago. Lo que importa, para el propósito que nos ocupa, es que las consideraciones éticas —que atañen a nuestra vida en sociedad con otras personas, y que incluyen, por ejemplo, las consideraciones relativas a la justicia y al respeto mutuo— entran a veces en conflicto con consideraciones relativas a un interés propio superficial[3].

Otro ejemplo que nos encontraremos en breve es el uso de «amoral» para describir a alguien que es completamente indiferente a consideraciones éticas.

[3] Esta expresión no quiere sugerir que las consideraciones relativas al interés propio sean *siempre* superficiales; véase el principio de la sección siguiente.

Esto significa que, si las consideraciones éticas se aducen como relevantes para la pregunta de Sócrates, se nos plantea la cuestión de cómo se justifican, que estructura los cinco capítulos siguientes del libro. Sin embargo, antes de pasar a esto, hay que tener claro qué cabe esperar de cualquier tipo de justificación. En concreto, afirma Williams, tenemos que tener claro:

- *contra* qué se da la justificación,
- *a* quién se da,
- *desde* dónde se da.

Con esto, Williams reacciona ante una clase de alarmismo que encuentra en gran parte de la filosofía moral. Un alarmismo que tiene su origen en dos factores: el primero es la convicción de que, si alguien es completamente amoral, esto es, si alguien no se ve motivado en absoluto por consideraciones éticas[4], debe ser posible ponerle remedio a esto esgrimiendo ante esa persona un *argumento* persuasivo idóneo, un argumento cuya elaboración constituye la tarea propia de la filosofía moral. El segundo factor que genera ese alarmismo es la pérdida de esperanzas en cuanto a la perspectiva de que la filosofía moral consiga elaborar un argumento de ese tipo. Williams comparte esta pérdida de esperanzas, pero no la convicción. En otras palabras, está de acuerdo en que no hay ninguna esperanza de que la filosofía moral elabore un argumento de esa clase, pero *no* está de acuerdo en que esta sea la tarea de la filosofía moral. Este es otro ejemplo de su escepticismo sobre el tipo de influencia que puede ejercer la filosofía. Compartir esa convicción —esto es, pensar que la tarea de la filosofía moral es elaborar un argumento como ese— supondría, en efecto, creer que debe haber una justificación de las consideraciones éticas que se puede dar *contra* el amoralismo, *al* amoralista y *desde* algún tipo de punto arquimédico, es decir, desde un conjunto de premisas que cabe esperar que el propio amoralista comparta.

Las esperanzas de Williams son más modestas; o, si no, son sin duda diferentes. Quiere buscar una justificación de las consideraciones éticas que pueda ofrecerse *contra* el amoralismo, pero no *al* amoralista, y, por lo tanto, no necesariamente *desde* un punto arquimédico. La

[4] Véase la nota 2.

justificación buscada es una que pueda ser planteada a aquellos para quienes las consideraciones éticas ya tienen cierta fuerza. En otras palabras, la cuestión no es *persuadir* a cualquiera de cualquier cosa, sino favorecer la autocomprensión, es decir, el tipo de autocomprensión que Williams entiende que es la verdadera tarea de la filosofía moral.

No «necesariamente» desde un punto arquimédico. Si no se espera que la justificación sirva como instrumento de conversión, entonces es evidente que no es tan necesario tratar de partir de premisas que comparta el amoralista. Pero, aun así, sí que hay *alguna* razón para ello. Cuanto más débiles sean los presupuestos sobre los que descansa la justificación, más profunda es la autocomprensión que esta puede propiciar.

Muy bien. Pero ¿hasta qué punto pueden ser débiles los presupuestos? ¿Es posible partir de un punto arquimédico? Williams no responde a esta pregunta en el capítulo 2. Lo que hace, ya al final del capítulo, es indicar dónde habría de residir el punto arquimédico, en caso de que *existiera:* «en la idea de acción racional» (pág. 54). Los dos siguientes capítulos exploran los dos intentos más conocidos y los dos mejores, en realidad, de partir de ese punto: el de Aristóteles, cuya concepción de la acción racional es relativamente rica y determinada; y el de Kant, cuya concepción de la acción racional es lo más mínima y abstracta posible. Si ninguno de los dos consigue su objetivo, el proyecto de justificar las consideraciones éticas desde un punto arquimédico o, como dice Williams, «de abajo arriba» (págs. 83 y 246), tendrá que ser abandonado.

Capítulo 3. Los fundamentos: el bienestar

Había una advertencia al inicio de la sección anterior: las consideraciones éticas pueden entrar en conflicto, a veces, con consideraciones relativas a un interés propio superficial. Para Aristóteles, «superficial» es la palabra clave. Que alguien actúe de acuerdo con consideraciones éticas consiste, en la concepción de Aristóteles, en hacer lo que *realmente,* o más *fundamentalmente,* supone su interés propio.

Hay diversas razones por las que alguien podría pensar que las consideraciones éticas y las relativas al interés propio coinciden en el fondo, y cualquiera de ellas podría invocarse para mostrar que es racional actuar de acuerdo con las primeras; en otras palabras, cualquiera de las razones que podrían invocarse en una justificación de las consideraciones éticas desde un punto arquimédico del tipo descrito al final de la

sección anterior. Por ejemplo, se podría afirmar que el castigo divino aguarda a quienes no actúan de acuerdo con consideraciones éticas. Aristóteles, sin embargo, entiende la conexión con la racionalidad de un modo más profundo, pues piensa que actuar de acuerdo con consideraciones éticas, o actuar virtuosamente, como él diría, es en sí mismo intrínsecamente racional, en cuanto que confiere la máxima coherencia posible a todo lo que uno está dispuesto a querer, a sentir o a hacer; y ello favorece el interés propio, porque el bienestar humano reside fundamentalmente en la vida racional que esencialmente distingue a los humanos de otros animales. (Así pues, hay un sentido en el que Aristóteles sostiene que actuar virtuosamente es tanto racional porque es en interés propio como que es en interés propio porque es racional).

Dado que Aristóteles concibe la principal tarea justificativa de la ética exactamente de la misma manera que Williams —aconsejar, por así decir, al convertido—, no tiene nada que decir a aquellos para quienes las consideraciones éticas carecen de toda fuerza. No tiene nada que decir*les*. Pero necesita decir algo *sobre* ellos. Necesita, como dice Williams, presentarnos «una *teoría del error*, una explicación sustantiva de la manera en la que las personas yerran en el reconocimiento de sus propios intereses auténticos» (pág. 72, cursivas añadidas). La explicación que les ofrece Aristóteles pivota en torno a la educación. Para Aristóteles, la virtud no puede conseguirse sin una formación adecuada, como también sucede con el resto de rasgos de la vida racional, sean las letras o los números. Quienes no reciben una formación adecuada en su infancia y adolescencia adquieren malos hábitos de búsqueda del placer que nublan su juicio.

Esta concepción no convence a Williams, principalmente porque no le convence la teleología subyacente que hace apropiado que nos planteemos en qué consiste, fundamentalmente, el bienestar humano. Y es también escéptico con respecto a que los desarrollos científicos modernos —en biología evolutiva o en psicología, pongamos por caso— puedan usarse para enmendar sus fallos. Ciertamente, Williams cree que el planteamiento aristotélico tiene algunos aciertos importantes, entre los que no es uno menor que las consideraciones éticas deriven la fuerza que puedan tener de la naturaleza humana, tal y como se manifiesta en las disposiciones de las personas. Pero sin la teleología subyacente, esto resulta insuficiente para fijar cuáles serán esas consideraciones. La naturaleza humana está sujeta a todo tipo de condicionamientos sociales e históricos, y se manifiesta en todo tipo de disposicio-

nes. Y estas disposiciones pueden dar apoyo a muchas perspectivas éticas diferentes, algunas de las cuales son mutuamente excluyentes. (Williams abundará en esto en el penúltimo capítulo, dedicado al relativismo). Retomando la cita anterior, no existe «la vida apropiada [...] para los seres humanos en cuanto tales».

Capítulo 4. Los fundamentos: la razón práctica

Rechazado el intento de Aristóteles de justificar las consideraciones éticas desde un punto arquimédico, Williams pasa a ocuparse de Kant, quien también quiere demostrar que es racional actuar de acuerdo a consideraciones éticas, o actuar por deber, en su terminología. Pero a diferencia de Aristóteles, Kant no concibe principalmente la cuestión en términos de bienestar humano, sino que parte de la idea de acción racional, prescindiendo de lo que los seres humanos, como especie o como individuos, pueden estar dispuestos a querer o a sentir. Kant defiende que estar motivado por consideraciones éticas es una precondición para ser un agente racional.

Williams no descarta que un argumento de este tipo pueda llegar a funcionar; específicamente, un argumento que concluya que una precondición de ser un agente racional es valorar la propia libertad. Pero esto está lejos de ser lo que Kant necesita. Valorar la propia libertad no es lo mismo que estar motivado por consideraciones éticas. (No es valorar la libertad de cualquier otro agente racional). ¿Cómo da Kant este paso adicional?

Lo da haciendo abstracción de *todo,* excepto de la agencia racional del agente. Kant cree que un agente racional, para ser fiel a su esencia, tiene que actuar por principios de la agencia racional *pura* («razón práctica pura»). Es decir, tiene que actuar sobre la base de principios que serían aptos para regular las acciones de *todos* los agentes racionales. Esto requiere que valore la libertad y la racionalidad, pero no su propia libertad, ni su propia racionalidad, sino la libertad y la racionalidad *per se*. Tiene que valorar a todos los seres humanos por sí mismos. Como dice Kant, un «ser racional tiene que considerarse siempre un legislador en un reino de los fines», entendiendo por «reino de los fines» una unión regida por leyes de los seres racionales, considerados como fines en sí mismos[5].

---

[5] Immanuel Kant, *Fundamentación de la metafísica de las costumbres,* Ak. IV, 434.

Actuar, para Kant, es en este respecto como pensar. Uno no piensa racionalmente a menos que piense de acuerdo con principios que serían aptos para regular el pensamiento de todos los sujetos pensantes racionales. Así, sería irracional pensar que el verdadero color de un objeto era cualquier color que a uno le pareciera que tenía a primera vista. Esto nos haría vulnerables a la posibilidad de que un objeto que a una persona primero le parecía amarillo, a otra persona a primera vista le pareciera naranja, quizá en diferentes condiciones lumínicas. (Su verdadero color no podría ser *tanto* amarillo *como* naranja. Tendría que haber algún principio sobre cuya base decidir entre esas apariencias incompatibles).

Es precisamente esta analogía entre actuar y pensar lo que Williams considera que es erróneo en el planteamiento de Kant. Actuar y pensar, para Williams, *no* son semejantes a este respecto. No se piensa racionalmente a menos que se piense de un modo conducente a creer la verdad, y la creencia en la verdad que tiene una persona es la misma que la creencia en la verdad de cualquier otra. Pero una persona puede actuar racionalmente de un modo que conduzca a la satisfacción de sus deseos, aunque lo que para ella satisface sus deseos puede ser bastante diferente de lo que para otra persona satisface los suyos, o incluso estar en conflicto. Para Williams, el intento de Kant de justificar las consideraciones éticas desde un punto arquimédico no tiene más éxito que el de Aristóteles.

Capítulo 5: Estilos de teoría ética

Ahora bien, estas conclusiones no niegan la posibilidad de que sea factible justificar las consideraciones éticas desde un planteamiento diferente al del punto arquimédico. Por ejemplo, teniendo en cuenta el tipo de fuerza que las consideraciones éticas pueden tener, puede ser posible justificar determinadas consideraciones éticas *específicas* contra sus rivales. Además, no hay razón alguna para rechazar que la justificación aristotélica y la justificación kantiana, cada una de las cuales puede ser errónea en sus propios términos, no puedan ser aprovechadas para elaborar una justificación de esta clase. (Así, aunque puede que no quepa un requisito racional como el que defendía Kant, el de aplicar a nuestra deliberación sobre cómo actuar la misma imparcialidad que aplicamos a nuestra deliberación sobre cómo pensar, sí puede ser un

requisito *ético)*. La forma más obvia que puede adoptar esa justificación es la de una *teoría ética*. En su esfuerzo por dilucidar si se puede lograr algo así, que le ocupa los capítulos 5 y 6, Williams tiene la oportunidad de examinar no solo la idea misma de teoría ética, sino también algunas de las teorías éticas concretas que se han propuesto, incluida una versión del utilitarismo, el cual, junto al aristotelismo y al kantismo, suele considerarse el tercer vértice del triángulo dialéctico que ha dominado la filosofía moral.

Una teoría ética, en la definición que da Williams, es «una caracterización teórica del pensamiento y la práctica éticos, que implica o bien un test general sobre la corrección de las creencias y los principios éticos básicos o, alternativamente, que no puede existir un test de este tipo» (pág. 104). La razón de esta definición disyuntiva, un tanto extraña, estriba en que da cuenta de los dos tipos de teoría que, sobre bases filosóficas, pretenden decirnos cómo hemos de pensar en ética. Pero consideremos caracterizaciones del segundo tipo (el que afirma que no puede existir un test sobre la corrección de las creencias y principios éticos básicos) y pensemos cuál sería el caso límite de tal caracterización. Sería la concepción según la cual «mantener una posición ética consiste simplemente en elegir una entre todas y atenerse a ella» (pág. 105). También esta concepción pretende decirnos, sobre bases filosóficas, cómo hemos de pensar en ética. Lo hace diciéndonos «que realmente no podemos pensar mucho en ética» (pág. 106).

Por el contrario, Williams quiere dar una caracterización del pensamiento y de la práctica éticos por medio de la cual podamos pensar en ética de todas las manera posibles, pero «la *filosofía* no puede decir gran cosa con respecto a cómo hemos de hacerlo» (pág. 106, cursiva añadida). Es tan escéptico con respecto a las expectativas de que se pueda formular una teoría ética sensata como acerca de las posibilidades de éxito de un proyecto de fundamentación del tipo emprendido por Aristóteles y por Kant.

Los dos estilos de teoría ética sobre los que dirige su mirada escéptica en el capítulo 5 son el contractualismo y el utilitarismo. El contractualismo es primo hermano del kantismo y sostiene que el pensamiento ético se ocupa de qué acuerdos informados y no forzados podrían alcanzar las personas. El utilitarismo mantiene que el pensamiento ético se ocupa del bienestar y de su maximización. Ambos dejan un amplio margen a un ulterior refinamiento (por ejemplo, en el caso del utilitarismo, sin precisar si son los actos individuales o las reglas o las prácticas o las instituciones lo que cabe evaluar en términos de la maximización

del bienestar, e incluso qué cuenta como bienestar). Las versiones del contractualismo y del utilitarismo que Williams examina son la de Rawls y la de Hare, respectivamente; versiones particularmente claras y convincentes, que, por lo tanto, suponen sólidas dianas ideales contra las que dirigir su desazón a propósito de ambos estilos de teoría.

Capítulo 6: Teoría y prejuicio

Volvamos a la idea de teoría ética que, como ya se ha indicado, constituye una de las dianas principales del libro.

¿Qué tipo de autoridad puede tener una teoría de este tipo? ¿A qué tiene que responder? En primer lugar, tiene que responder a las intuiciones que tenemos (por ejemplo, sobre qué sería o no aceptable que hagamos en diversas situaciones). Esto no significa negar que una teoría ética pueda usarse para criticar y reemplazar algunas de nuestras intuiciones. De hecho, una de las funciones que se espera desempeñe esta teoría es precisamente la de eliminar conflictos entre nuestras intuiciones, usando unas para invalidar otras. No obstante, el caso es que ninguna teoría ética puede desempeñar esta función si no es mediante la imposición de alguna estructura coherente y manejable sobre nuestras intuiciones, que preserve tantas intuiciones como sea posible.

Ninguna *teoría ética* puede desempeñar esta función si no es de esta manera. Hay otros modos, menos sistemáticos, de eliminar conflictos entre nuestras intuiciones. Por ejemplo, podemos sencillamente ejercer nuestro juicio acerca de cada conflicto particular que se plantee. Las teorías éticas no pueden reclamar una autoridad especial solo por su capacidad para eliminar conflictos. ¿De dónde, pues, deriva su supuesta autoridad? En gran medida, de lo que Williams denomina «una concepción racionalista de la racionalidad» (pág. 41). Se trata de una aplicación a la deliberación personal de un ideal de vida pública que exige que, «en principio, toda decisión se base en fundamentos que se puedan explicar de forma argumentada» (pág. 41); un ideal que no se realiza solo con tomar una decisión al juzgar en algún caso particular. Pero ¿por qué hemos de aceptar la aplicación de este ideal a la deliberación personal? ¿No nos incita esto a buscar un orden, una sistematicidad y una economía de ideas que no son idóneas para las complejidades de la deliberación personal en la vida real? Y, en todo caso, ¿qué añade el ideal a las intuiciones mismas? Como Williams insiste en otro sitio,

«"No puedes matarlo, es un niño" es más convincente como razón que cualquier otra que pueda aducirse por ser una razón»[6] (cfr. págs. 186-187).

A buen seguro, es importante que reflexionemos sobre nuestras intuiciones. Al hacerlo, podemos descubrir que algunas son prejuicios irracionales, pero son irracionales en tanto que se basan en el autoengaño o en el engaño social, digamos, no porque entren en conflicto con alguna teoría ética que hayamos ideado. No es un requisito de la reflexión que dé lugar a ninguna clase de teoría. Ni, por ello mismo, ha de darse ninguna importancia especial a aquellas de nuestras ideas éticas que son el producto de la reflexión.

Los utilitaristas, como es bien sabido, le dan una importancia especial a las ideas éticas que resultan de la reflexión; en concreto, debido a un rasgo tan habitual como objetable de su teoría como es el de *promover* la disonancia entre aquellas de nuestras ideas éticas que son el producto de la reflexión y las que no lo son. (En su forma menos objetable, el contraste se establece entre diferentes ideas que tenemos en diferentes momentos: en la hora serena de la reflexión y en el calor del momento. En su versión más objetable, el contraste se da entre diferentes ideas que tienen diferentes *grupos:* la élite reflexiva y el resto. Esto último es lo que Williams denomina el «utilitarismo elitista», pág. 144). El utilitarismo tiene esta característica porque las intuiciones a su favor, que sus defensores ven como el producto de la reflexión ilustrada, proporcionan una razón para preservar y alentar el pensamiento no utilitarista al nivel irreflexivo: esto se debe a que es más probable que las personas maximicen el bienestar a ese nivel si tratan de hacer algo que no sea maximizar el bienestar.

Al final del capítulo 6, la idea de una teoría ética queda más o menos debilitada a la luz del escepticismo general de Williams acerca de la ética filosófica; «un escepticismo», comenta en la página 106, «referido a la filosofía más que a la ética».

Capítulo 7: El giro lingüístico

La filosofía moral debate cuestiones de gran envergadura concernientes a la *metafísica* del valor. Por ejemplo, ¿hay alguna distinción fundamental entre hechos y valores, entre cómo son las cosas con inde-

---

[6] Bernard Williams, «Conflicto de valores», reimpreso en *La fortuna moral.*

pendencia de lo que nosotros pensamos sobre ellas y las evaluaciones que proyectamos en relación con cómo son las cosas? Hasta el momento, estas cuestiones han permanecido en un segundo plano. En los capítulos 7, 8 y 9, Williams las aborda explícitamente. Su interés en el capítulo 7 es examinar la cuestión de en qué medida podemos mejorar nuestra comprensión de estos asuntos usando la principal herramienta metodológica de la filosofía analítica: el análisis del lenguaje.

Muchas personas creen que existe una distinción entre palabras evaluativas como «atroz», «supererogatorio», «reprensible» y «bueno», y palabras no evaluativas, tales como «sulfúrico», «octogenario», «impermeable» y «rubio», y que es imposible definir una palabra del primer tipo usando solo palabras del segundo tipo. El nombre «falacia naturalista», acuñado por Moore[7], suele usarse para designar el intento erróneo de hacer esto mismo. (Pero, como dice Williams: «Es difícil encontrar en la historia de la filosofía un nombre tan espectacularmente erróneo para una idea», porque normalmente se entiende que una «falacia» es un error en la inferencia, y que «naturalista» remite a una concepción según la cual «la ética había de entenderse en términos seculares, sin ninguna referencia a Dios o a cualquier otra autoridad trascendente» [págs. 158-159]; pero nada de esto tiene mucho que ver con el intento de definir palabras evaluativas usando solo palabras no evaluativas). Siempre y cuando sea adecuado trazar esa distinción, es razonable esperar que nuestra comprensión de la metafísica del valor mejore mucho si prestamos atención a cómo usamos las palabras de cada uno de estos dos tipos.

Sin embargo, para Williams es justo al revés. En tanto que tenemos alguna idea de a qué se supone que prestamos atención, y en tanto que pueda establecerse esa distinción lingüística, ello se debe a la comprensión con la que nos aproximamos al lenguaje relativo a la distinción metafísica entre hechos y valores. «En tanto que» es, en cualquier caso, la parte crucial. Porque, aunque el propio Williams reconoce una distinción de esta clase entre hechos y valores, se trata de una distinción muy sutil, que cree que nuestro lenguaje no refleja muy bien. Por el contrario, considera que nuestro lenguaje contribuye a ocultárnosla, y a fomentar diversas ilusiones sobre la metafísica del valor (y sobre la naturaleza de la ética en general).

---

[7] G. E. Moore, *Principia Ethica,* sec. 10.

Lo que realmente encontramos en el lenguaje son cientos y cientos de palabras «híbridas», tales como «casta», «desleal», «brutal» y «orgulloso». Son palabras que representan lo que Williams denomina concepto éticos «densos». La noción de concepto ético denso es extremadamente importante para él. Un concepto ético denso es un concepto que tiene *tanto* un aspecto evaluativo, pues aplicarlo a una situación dada supone, en parte, evaluar la situación, *como* un aspecto fáctico, puesto que aplicarlo a una situación dada es emitir un juicio que está sujeto a corrección si la situación resulta que no es así. De este modo, si afirmo que una persona ha sido desleal, la estoy censurando; pero también digo algo claramente falso si resulta que no ha traicionado ningún acuerdo relevante. El concepto de deslealtad no es solo un concepto evaluativo con un indicador de la desaprobación. Williams, a diferencia de muchos de los que se han ocupado de estos conceptos, defiende enérgicamente que hechos y valores están *inextricablemente* entrelazados. Esta es una razón de por qué el lenguaje en el que se formulan proporciona una indicación deficiente de la metafísica subyacente.

Así pues, el análisis del lenguaje tiene un uso muy limitado en filosofía moral. Sin embargo, sí tiene cierta utilidad. Puede servir para recordarnos que nuestra vida ética, igual que nuestro lenguaje ético, es un complejo fenómeno social multifacético, que varía con el tiempo y de un grupo a otro; y la comprensión ética, que necesita dar cuenta de esa variación, también «requiere una dimensión explicativa de tipo social» (pág. 168).

Capítulo 8: Conocimiento, ciencia y convergencia

El capítulo 8 es el núcleo del libro. Este es el capítulo en el que Williams afronta directamente los asuntos relativos a la metafísica del valor, esto es, la cuestión de si existe alguna distinción fundamental entre hechos y valores, y similares.

En cierto sentido, son también asuntos sobre la *objetividad* del pensamiento ético, y desde este planteamiento los aborda Williams, pues entiende que hay un tipo de objetividad que, para toda concepción relativista de la cuestión, no es aplicable a nuestro pensamiento ético, aunque sea aplicable a nuestro pensamiento en otras áreas. (Esto conecta con la tesis mencionada anteriormente de que Williams reconoce una distinción de esta clase entre hechos y valores). La cuestión es: ¿pero *qué* tipo de objetividad?

La palabra «objetividad» tiene usos muy distintos. Pero, en todas sus interpretaciones, guarda alguna relación con el acuerdo. Decir que existe un tipo de objetividad que se aplica o no a nuestro pensamiento en un área dada es decir algo acerca de la posibilidad de alcanzar un acuerdo regido por principios en esa área o, como dice Williams, de que *converjamos* en nuestras creencias en esa área. Muy bien, pues; ¿qué es exactamente lo que Williams está dispuesto a decir sobre qué opciones hay de converger en algunas de nuestras creencias que no está dispuesto a decir sobre las opciones que hay de que converjamos en nuestras creencias éticas? Esta resulta ser una cuestión extraordinariamente delicada.

La posición de Williams no es que sea razonable esperar converger en algunas de nuestras creencias, pero que no es razonable que esperemos hacerlo en nuestras creencias éticas. Menos aún es que converjamos, de hecho, en algunas de nuestras creencias, pero que nunca lo hagamos en nuestras creencias éticas. Ni tampoco tiene que ver con si, donde *hay* convergencia, las creencias en cuestión merecen o no el título de «conocimiento». Tiene que ver con las diferentes maneras de *explicar* cualquier convergencia existente. El contraste fundamental es entre ciencia y ética.

La posición de Williams es la siguiente: a veces, convergemos en nuestras creencias éticas, y esas creencias merecen, a veces, el título de «conocimiento». Esto puede suceder cuando hay un concepto ético denso involucrado en las creencias en cuestión. Así pues, las personas que usan el concepto de castidad podrían no tener dificultades en estar de acuerdo sobre si cierto acto es casto y, de hecho, podrían saberlo. Sin embargo, el quid de la cuestión reside, en primer lugar, en lo que está implicado en la manera en que usan el concepto de castidad. Asumida la combinación distintiva de evaluación y factualidad propia del concepto, su uso es parte de la vida en un mundo social particular, un mundo en el que se aprecian determinadas cosas y se abomina de otras. Las personas necesitan vivir en *un* mundo social. Pero, como demuestra la historia sobradamente, no hay un mundo social determinado en el que las personas necesiten vivir. No necesitan vivir en un mundo que respalde el concepto de castidad. Así pues, cualquier buena explicación reflexiva de por qué las personas convergen en sus creencias sobre qué es casto tiene que incluir una explicación de por qué usan el concepto de castidad, de por qué viven en *ese* mundo social. (Esta es la «dimensión explicativa de tipo social» a la que Williams se refiere al final del

capítulo 7). Esta explicación no puede invocar el concepto de castidad, porque tiene que ser una forma de reflexión desde una posición privilegiada externa al mundo social en cuestión. Por ello, no puede respaldar directamente sus creencias. (Esto es, no puede conformarse al esquema: «Esas personas convergen en sus creencias acerca de *x* porque son apropiadamente sensibles a las verdades acerca de *x*». No puede representarlas como concordando en qué es casto por los conocimientos que tienen de qué es casto). Por el contrario, una buena explicación reflexiva de por qué las personas convergen en sus creencias sobre un conjunto particular de cuestiones científicas, en sus creencias sobre cómo es el oxígeno, por ejemplo, puede invocar los mismos conceptos que están funcionando en las creencias y, por lo tanto, si las creencias se han adquirido adecuadamente, pueden respaldarlas. (Esto *puede* ajustarse al esquema específico anterior. Se puede representar a esas personas como concordando sobre cómo es el oxígeno por los conocimientos que han adquirido de cómo es el oxígeno, por lo que han descubierto sobre el oxígeno).

Una consecuencia de esta posición es que, sea el que sea el conocimiento ético que posean las personas, lo tienen por aplicar sin titubeos y de forma espontánea sus conceptos éticos densos. No hay conocimiento ético al que se llegue por reflexión sobre si es o no «correcto» usar esos conceptos. Esta es la razón por la que Williams presenta su argumento a favor de la existencia del conocimiento ético invocando la ficción de una sociedad «hipertradicional», esto es, «una sociedad homogénea en el grado máximo y con una tendencia mínima a la reflexión general» (pág. 180). Es ahí donde, para Williams, se hallan los ejemplos más claros de conocimiento ético.

Pero Williams va más allá. Defiende que en una sociedad como la nuestra, en la que abunda la reflexión, la reflexión puede tener un efecto perturbador. Las personas pueden llegar a abandonar algunos de sus conceptos éticos densos porque se dan cuenta de que están vinculados a creencias falsas, por ejemplo, o simplemente porque son conscientes de la existencia de alternativas. Eso les impide retener cualquier conocimiento que tuvieran por manejar esos conceptos. Es así como Williams extrae una de las conclusiones más chocantes y controvertidas del libro: «la conclusión, destacadamente antisocrática», como la llama, «de que, en ética, *la reflexión puede destruir el conocimiento* (pág. 187, la cursiva es suya). Esta conclusión es «antisocrática» porque Sócrates, cuya pregunta reflexiva inició toda esta investigación, insistía en que

una vida sin reflexión —«una vida sin examen»— no valía la pena ser vivida, creía que «no puede haber conocimiento sin reflexión» (pág. 208).

Capítulo 9: Relativismo y reflexión

El contraste entre la ciencia y la ética que Williams explora en el capítulo 8 le lleva a decir que «la ciencia tiene buenas opciones de ser más o menos lo que parece, una exposición teórica sistemática de cómo es el mundo, mientras que el pensamiento ético no tienen ninguna opción de ser todo lo que parece» (pág. 172). En concreto, el pensamiento ético «no puede poner de manifiesto que descansa en disposiciones humanas» (pág. 244). Hice referencia, en la sección inicial, a la esperanza que expresa Williams en el Epílogo de que nuestra experiencia ética pueda aguantar el embate de cualquier autocomprensión que la desvele como algo distinto a lo que parece. En el capítulo 9, Williams se ocupa de la pregunta de cómo puede ser eso, dado el ataque del capítulo 8.

Lo que nos hace falta, afirma, es *confianza*. Y este es un fenómeno social. Aunque son los individuos los que confían, su confianza normalmente se promueve y se refuerza mediante mecanismos sociales como la educación, el apoyo institucional y el discurso público. (Lo que no ayuda mucho, insiste Williams, desarrollando uno de sus temas principales, es la filosofía, más bien, esta contribuye a que la necesitemos). La confianza permite a los individuos atenerse a sus conceptos éticos densos con independencia de los efectos perturbadores de la reflexión. La confianza es buena, pero no supremamente buena. Algunas maneras de lograrla, por ejemplo suprimiendo la argumentación racional, conllevan un sacrificio indebido de otros bienes, y hay que oponerse a ellas.

Otra cuestión que aborda Williams en el capítulo 9 es la de qué forma de relativismo, si la hubiese, implica esta concepción; es decir, esta concepción de mundos sociales diferentes que sostienen diferentes conceptos éticos densos, en algunos casos hasta el punto de ser irreconciliables. No se trata de la forma más cruda de relativismo, afirma Williams, según la cual habríamos de tener «la misma buena disposición hacia las creencias éticas de todo el mundo» (pág. 198). No hay nada en esta concepción que nos impida despreciar las creencias éticas de ciertas personas, ni intentar combatirlas cuando esas creencias nos afectan. ¿Cómo es posible? Lo que es, señala Williams, «gravemente

erróneo» es pensar que un relativismo de las creencias éticas puede dar lugar a «una moral no relativista de la tolerancia universal» (pág. 198). Incluso así, la concepción de Williams, al dirigir nuestra atención a las diferencias más llamativas entre nuestra perspectiva ética y las de otras sociedades, está destinada a dejarnos insatisfechos y perplejos, «Nosotros estamos en lo cierto y el resto del mundo está equivocado». ¿No implica esto *algún* tipo de relativismo?

Hablando en sentido estricto, Williams cree que no; esto es, que no *impide* que nos sintamos perplejos ante el pensamiento «Nosotros estamos en lo cierto, y el resto del mundo está equivocado». Sin embargo, al haber mostrado que es una posición poco atractiva, *deja margen* a alguna forma de relativismo, a alguna manera de ir más allá de la perplejidad. Es en este punto en el que Williams introduce lo que denomina «el relativismo de la distancia» (pág. 201), según el cual solo una sociedad lo suficientemente «cercana» a la nuestra, esto es, solo cuando existe la opción real de que adoptemos la perspectiva ética de esa sociedad, se plantea la cuestión de *valorar* esa perspectiva ética como «correcta», «incorrecta», «injusta», etc. Williams cree que su concepción propicia una versión matizada del relativismo; «matizada» porque no niega que se pueda realizar *cierto tipo* de evaluación de las perspectivas éticas de sociedades distantes y puede incluso ser necesaria en el caso específico de la evaluación con respecto a la justicia. Este relativismo de la distancia matizado puede parecer muy atenuado, pero se plantea, de nuevo, el contraste con la ciencia: una perspectiva científica, por distante que sea la sociedad a la que pertenece, tiene que considerarse siempre correcta o incorrecta.

Capítulo 10: La peculiar institución de la moral

El capítulo 10 es una especie de apéndice al resto del libro. En el capítulo 1, como ya vimos, Williams da una indicación temprana de su antipatía por el estilo particular de pensamiento ético que llama «la moral». En el capítulo 10, explica qué es la moral, «y por qué estaríamos mejor sin ella» (pág. 215).

Hay dos observaciones que conviene añadir a lo que se ha dicho ya sobre esto: en primer lugar, aunque Williams se opone a la idea de obligación moral, no se opone a *todo* lo relativo a la obligación. Admite sin duda que, para vivir juntos en sociedad, hace falta que tengamos deter-

minadas expectativas básicas y más o menos categóricas (como la expectativa de que no nos mientan, o de que no nos maten). Además, la única manera en que puede servir de ayuda una vida ética es consiguiendo inculcar en las personas disposiciones a tratar los requisitos correspondientes (en los casos anteriores, de no mentir y de no matar) como obligaciones. Alguien que esté sujeto a esa obligación puede concluir que le resulta *imposible* hacer cierta cosa, o que *tiene que* hacerla. Pero, insiste Williams, este no es un tipo de conclusión peculiar de la ética, en contraposición con la moral. Podemos llegar al mismo tipo de conclusión «por razones de prudencia, de autoprotección, de interés estético o artístico o de pura autoafirmación» (pág. 231).

La segunda observación es que Williams da un diagnóstico muy persuasivo de por qué se apela a la moral: expresa «el ideal de que la existencia humana pueda ser en último término justa» (pág. 239), y lo hace concibiendo la cualidad personal que importa más que ninguna otra, a saber ser moral, como una cualidad que escapa a los vaivenes de la suerte, en contraste con ser feliz o tener talento o ser amado, pongamos por caso. Pero es aquí donde estriba la mayor discrepancia de Williams con la moral: «la idea de un valor que escapa a la suerte», insiste, «es una ilusión» (pág. 240). Es la idea de un valor que está «más allá de cualquier determinación empírica»; un valor que reside «solo en el hecho de intentarlo y no en el de conseguirlo, puesto que conseguir algo depende en parte de la suerte, pero tiene que ser un tipo de intento que se sitúe más allá de la medida en la que la capacidad misma de intentar puede ser cuestión de suerte» (pág. 239). No hay, para Williams, ningún lugar en el que pueda situarse. Concluye el capítulo reprendiendo a la moral por ser «una concepción equivocada de la vida que está profundamente enraizada y es aún poderosa» (pág. 240).

Lecturas complementarias

Las obras de Williams que están más relacionadas con *La ética y los límites de la filosofía* (en adelante, ELF) son: *Morality: An Introduction to Ethics* (Cambridge, Canto, 1993) [*Introducción a la ética,* trad. de M. Jiménez Redondo (Madrid, Cátedra, 1988)], una introducción maravillosamente concisa a la disciplina que funciona como compendio de muchas de las ideas principales de ELF; *Utilitarianism: For and*

Against (Cambridge, Cambridge University Press, 1973) [*Utilitarismo. Pro y contra,* trad. de J. Rodríguez Marín (Madrid, Tecnos, 1981)], escrito junto con J. J. C. Smart, donde la contribución de Williams, «A Critique of Utilitarianism», presenta los argumentos en contra del utilitarismo; y *Shame and Necessity* (Berkeley y Los Ángeles, University of California Press, 1993) [*Vergüenza y necesidad,* trad. de A. Montes Sánchez (Madrid, A. Machado, 2011)], libro en el que desarrolla su interés por el pensamiento ético griego. También son relevantes *Descartes: the Project of Pure Enquiry* (Londres, Routledge, 2005) [*Descartes: el proyecto de una investigación pura,* trad. de J. A. Coll (Madrid, Cátedra, 1996)], en el que proporciona algunas de las herramientas básicas para elaborar el contraste fundamental que plantea entre el pensamiento ético y el pensamiento científico; y *Truth and Truthfulness: An Essay in Genealogy* (Princeton, Princeton University Press, 2002) [*Verdad y veracidad: una aproximación genealógica,* trad. de A. E. Álvarez y R. Orsí (Barcelona, Tusquets, 2006)], donde ofrece una caracterización nietzscheana de las virtudes de la precisión y de la sinceridad, que desarrolla algunas de las esperanzas que expresa en el Epílogo a ELF.

Los ensayos de Williams más relevantes para ELF son: el último de los seis artículos que aparecen en *Problems of the Self: Philosophical Papers 1956-1972* (Cambridge, Cambridge University Press, 1973) [*Problemas del yo,* trad. de J. M. G. Holguera (México, UNAM, 1984; 2.ª ed. rev., 2013)]; todos, salvo los dos últimos, de *Moral Luck: Philosophical Papers 1973-1980* (Cambridge, Cambridge University Press, 1981) [*La fortuna moral,* trad. de S. Marín (México, UNAM, 1993)]*; todos los que constituye *Making Sense of Humanity and Other Philosophical Papers 1982-1993* (Cambridge, Cambridge University Press), en especial los de las partes I y III; «Truth in Ethics», en Brad Hooker (ed.), *Truth in Ethics* (Oxford, Basil Blackwell, 1996); todos los artículos de *In the Beginning Was the Deed: Realism and Moralism in Political Argument*, editado por Geoffrey Hawthorn (Princeton, Princeton Uni-

---

* En castellano puede verse también: B. Williams y T. Nagel, *La suerte moral,* traducción e introducción de S. Rosell (Oviedo, KRK, 2013), que recoge las aportaciones de ambos filósofos a esta importante cuestión, que ellos mismos introdujeron en la agenda filosófica. Se incluye también la «Posdata» de Williams, publicada originalmente en *Making Sense of Humanity,* en la que clarifica su posición al respecto ante ciertos malentendidos. *[N. del T.].*

versity Press, 2005) [*En el principio era la acción,* trad. de A. García de la Serna (México, FCE, 2012)]; y todos los de la parte II de *Philosophy as a Humanistic Discipline,* editado por A. W. Moore (Princeton, Princeton University Press, 2006) [*La filosofía como disciplina humanística,* trad. de A. García de la Serna (México, FCE, 2011)]. Cada uno de estos ensayos explora, con mayor o menor profundidad, alguna cuestión particular, o conjunto de ellas, de la filosofía moral.

Hay muchas reseñas de ELF. Las dos más destacadas son la contribución de Simon Blackburn a «Making Ends Meet: A Discussion of *Ethics and the Limits of Philosophy*», *Philosophical Books,* 27 (1986), a la que responde Williams; y John McDowell, «Critical Notice of *Ethics and the Limits of Philosophy*», *Mind,* 95 (1986).

Una compilación excelente de artículos sobre la filosofía moral de Williams, muy inspirada en ELF, es la de J. E. J. Altham y Ross Harrison (eds.), *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams* (Cambridge, Cambridge University Press, 1995). En ella merecen mención especial: John McDowell, «Might There Be External Reasons?», que examina el proyecto de fundamentar la ética en la razón pura; Martha Nussbaum, «Aristotle on Human Nature and the Foundations of Ethics», que se ocupa del proyecto fundacional de Aristóteles; Christopher Hookway, «Fallibilism and Objectivity: Science and Ethics», y Nicholas Jardine, «Science, Ethics, and Objectivity», que se interesan por las distinciones que Williams traza entre ciencia y ética; J. E. J. Altham, «Reflection and Confidence», que examina la tesis de que la reflexión puede destruir el conocimiento; y Charles Taylor, «A Most Peculiar Institution», que aborda el tratamiento de «la moral» de Williams. Hay réplica de Williams a todos ellos.

Otra excelente compilación que se inspira principalmente en ELF es la de Edward Harcourt (ed.), *Morality, Reflection, and Ideology* (Oxford, Oxford University Press, 2000). Particularmente recomendables son la introducción del propio Harcourt y el ensayo de Miranda Fricker, «Confidence and Irony», que explora la noción de confianza de Williams. El volumen también contiene otro excelente escrito de Williams, «Naturalism and Genealogy».

Para un examen crítico de la concepción de la ciencia de Williams, véase Hilary Putnam, *Renewing Philosophy* (Cambridge, Mass., Harvard University Press, 1992) [*Cómo renovar la filosofía,* trad. de C. Laguna (Madrid, Cátedra, 1994)], capítulo 5, «Bernard Williams and the Absolute Conception of the World». Para un escrutinio ulterior de la idea

de que la reflexión puede destruir el conocimiento, véase A. W. Moore, «Williams on Ethics, Knowledge, and Reflection», *Philosophy*, 78 (2003); y Warren Quinn, «Reflection and the Loss of Moral Knowledge: Williams on Objectivity», en su libro *Morality and Action* (Cambridge, Cambridge University Press, 1993).

Bernard Williams fue uno de los más grandes filósofos de su generación. *La ética y los límites de la filosofía* no solo ha sido reconocida como su obra más importante, sino que se ha convertido también en un clásico contemporáneo de la filosofía moral, marcando los principales temas de debate. No se puede entrar en el mundo del pensamiento ético contemporáneo sin tomar posición con respecto a las cuestiones planteadas por Williams.

La ética y los límites de la filosofía nos propone que cambiemos la manera en la que hacemos ética. Bernard Williams sostiene que los filósofos no deberían seguir dando apoyo a lo que él llama «el sistema de la moral» —una estructura punitiva que pivota sobre la obligación, la censura y la culpa—, que se sustenta en la construcción de teorías morales, de la obligación, del castigo justo, etc., cada vez más complejas. Deberíamos centrarnos, en cambio, en esta otra cuestión: en qué consistiría vivir nuestras vidas éticas con confianza, de manera legítima.

El libro también ha tenido un efecto perturbador en el estudio académico de la ética, en tanto que reta a profesores y estudiantes a que reflexionen sobre el tipo de actividad en que están inmersos y cuestionen su significado.

Bernard Williams explora y reflexiona acerca de los problemas más complejos de la filosofía contemporánea y plantea nuevas ideas sobre temas centrales como el relativismo, el objetivismo y la posibilidad de un conocimiento ético.

Esta edición incluye comentarios al texto a cargo del filósofo A. W. Moore.

IMPORTADO POR
LA FAMILIA
DISTRIBUIDORA DE LIBROS S.A.
R.U.C. 20100365856
Av. Repúblicas de Chile N° 661
JESUS MARIA

0112103

ISBN 978-84-376-3561-3

colección TEOREMA
serie mayor